

10/526572

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
18. März 2004 (18.03.2004)

PCT

(10) Internationale Veröffentlichungsnummer
WO 2004/022586 A3

(51) Internationale Patentklassifikation⁷: C12P 1/04,
C07K 5/078, C12R 1/01, A61K 38/05, C12N 1/20

(74) Anwälte: BOETERS, Hans, D. usw.; Boeters & Lieck,
Bereiteranger 15, 81541 München (DE).

(21) Internationales Aktenzeichen: PCT/EP2003/009780

(81) Bestimmungsstaaten (*national*): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,
SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(22) Internationales Anmeldedatum:
3. September 2003 (03.09.2003)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
102 41 152.2 5. September 2002 (05.09.2002) DE

(71) Anmelder (*für alle Bestimmungsstaaten mit Ausnahme
von US*): GESELLSCHAFT FÜR BIOTECHNOL-
OGISCHE FORSCHUNG MBH (GBF) [DE/DE];
Mascheroder Weg 1, 38124 Braunschweig (DE).

(84) Bestimmungsstaaten (*regional*): ARIPO Patent (GH,
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL,
PT, RO, SE, SI, SK, TR), OAPI Patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht

(72) Erfinder; und

(75) Erfinder/Anmelder (*nur für US*): HOEFLE, Gerhard
[DE/DE]; Mascheroder Weg 1, 38124 Braunschweig
(DE). MUELLER, Rolf [DE/DE]; Mascheroder Weg 1,
38124 Braunschweig (DE). SASSE, Florenz [DE/DE];
Mascheroder Weg 1, 38124 Braunschweig (DE). SAND-
MANN, Axel [DE/DE]; Mascheroder Weg 1, 38124
Braunschweig (DE). BLOECKER, Helmut [DE/DE];
Mascheroder Weg 1, 38124 Braunschweig (DE).

(88) Veröffentlichungsdatum des internationalen
Recherchenberichts: 11. November 2004

Zur Erklärung der Zweibuchstaben-Codes und der anderen Ab-
kürzungen wird auf die Erklärungen ("Guidance Notes on Co-
des and Abbreviations") am Anfang jeder regulären Ausgabe der
PCT-Gazette verwiesen.

(54) Title: TUBULYSIN BIOSYNTHESIS GENE

(54) Bezeichnung: TUBULYSIN-BIOSYNTHESE-GENE

(57) Abstract: The invention relates to an ss-DNA molecule, a ds-DNA molecule, an RNA, a vector, the use of the same, and a cell,
especially for expressing tubulysin genes.

(57) Zusammenfassung: Die Erfindung betrifft ein ss-DNA-Molekül, ein ds-DNA-Molekül, eine RNA, einen Vektor, dessen Ver-
wendung sowie eine Zelle, insbesondere zur Expression von Tubulysin-Genen.

Express Mail No.
EV195875530US

WO 2004/022586 A3

Gesellschaft für Biotechnologische Forschung mbH

Tubulysin-Biosynthese-Gene

Tubulysine sind bereits als eine neue, auf das Tubulin-Skelett wirkende Substanzfamilie aus Myxobakterien in Irsee vorgestellt worden; vgl. PCT/EP 97/05095 und DE 100 08 089.8 und die darin angeführte Literatur. Im Gegensatz zu den Epothilonen zeigen diese eine mikrotubuli-abbauende Wirkung sowie die vermehrte Ausbildung von Zentrosomen. Mit einer Cytotoxizität von $IC_{50} = 10 - 500$ pg sind die Tubulysine als potentielle Cytostatika von besonderem Interesse.

Die Tubulysine haben eine cytostatische oder antimitotische Wirkung auf Pilze, humane Tumore oder Krebszelllinien und andere tierische Zellkulturen (vgl. Tabelle). Sie führen in den Zellen zu einem raschen Abbau des Mikrotubuli-Gerüsts. Das Aktinskelett bleibt erhalten. Adhärent wachsende L929-Maus-Zellen vergrößern unter dem Einfluß der Tubulysine ihr Zellvolumen, ohne sich zu teilen, und entwickeln große Zellkerne, die dann in einem apoptischen Vorgang zerfallen.

Wirkunsspektrum

<u>Pilze</u>	<u>Hemmhof [mm]</u>	
	<u>Tubulysin A</u>	<u>Tubulysin B</u>
<i>Aspergillus niger</i>	20	18
<i>Botrytis cineria</i>	23	18
<i>Coprinus cinereus</i>	20	
<i>Pythinum debaryanum</i>	20	

Agardiffusionstest: 20 µg pro Testblättchen 6 mm Durchmesser

<u>Humane Krebszellinien</u>	<u>IC₅₀ [ng/ml]</u>		
	<u>Tubulysin A</u>	<u>Tubulysin B</u>	<u>Tubulysin C</u>
KB-3-1 (DSM ACC 158)	0,01	0,02	0,1
K-562 (ATCC CCL 243)	0,1	0,2	1,5
HL-60 (ATCC-CCL 240)	0,04	0,08	0,4

Tierische Zellinien

L929, Maus (ATCC CCL1)	0,2	0,4	2
Pt K2, <i>Potorous tri-</i> <i>dactylis</i> (ATCC CCL 56)	0,2	0,2	2

Gemäß einer Ausführungsform betrifft die Erfindung ein ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:

- (i) ss-DNA-Molekül mit einer Sequenz gemäß Figur 1;
- (ii) ss-DNA-Molekül, das mit einem ss-DNA-Molekül gemäß (i), hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zu jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von dem ss-DNA-Molekül gemäß (i) hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und
- (iii) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines ss-DNA-Moleküls gemäß (i) oder (ii) komplementär ist.

Ferner betrifft die Erfindung ein ds-DNA-Molekül aus einem erfindungsgemäßen ss-DNA-Molekül und einem dazu komplementären Strang.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung ein ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:

- (i) ss-DNA-Molekül mit einer Sequenz der Positionen 3.308 bis 1 (ORF 16) der Sequenz gemäß Figur 1;
- (ii) ss-DNA-Molekül mit einer Sequenz der Positionen 4706 bis 3453 (ORF 15) der Sequenz gemäß Figur 1;
- (iii) ss-DNA-Molekül mit einer Sequenz der Positionen 5719 bis 7164 (ORF 14) der Sequenz gemäß Figur 1;

- (iv) ss-DNA-Molekül mit einer Sequenz der Positionen 9557 bis 7317 (ORF 13) der Sequenz gemäß Figur 1;
- (v) ss-DNA-Molekül mit einer Sequenz der Positionen 12193 bis 10550 (ORF 12) der Sequenz gemäß Figur 1;
- (vi) ss-DNA-Molekül mit einer Sequenz der Positionen 12841 bis 13881 (ORF 11) der Sequenz gemäß Figur 1;
- (vii) ss-DNA-Molekül mit einer Sequenz der Positionen 14833 bis 13835 (ORF 10) der Sequenz gemäß Figur 1;
- (viii) ss-DNA-Molekül mit einer Sequenz der Positionen 14942 bis 15586 (ORF 9) der Sequenz gemäß Figur 1;
- (ix) ss-DNA-Molekül mit einer Sequenz der Positionen 15847 bis 16983 (ORF 8) der Sequenz gemäß Figur 1;
- (x) ss-DNA-Molekül mit einer Sequenz der Positionen 21154 bis 18809 (ORF 7) der Sequenz gemäß Figur 1;
- (xi) ss-DNA-Molekül mit einer Sequenz der Positionen 22366 bis 23532 (ORF 6) der Sequenz gemäß Figur 1;
- (xii) ss-DNA-Molekül mit einer Sequenz der Positionen 24591 bis 26513 (ORF 5) der Sequenz gemäß Figur 1;
- (xiii) ss-DNA-Molekül mit einer Sequenz der Positionen 26597 bis 27517 (ORF 4) der Sequenz gemäß Figur 1;

- (xiv) ss-DNA-Molekül mit einer Sequenz der Positionen 29858 bis 30400 (ORF 3) der Sequenz gemäß Figur 1;
- (xv) ss-DNA-Molekül mit einer Sequenz der Positionen 31220 bis 32392 (TubA) der Sequenz gemäß Figur 1;
- (xvi) ss-DNA-Molekül mit einer Sequenz der Positionen 33056 bis 32397 (ORF 2) der Sequenz gemäß Figur 1;
- (xvii) ss-DNA-Molekül mit einer Sequenz der Positionen 34195 bis 33074 (TubZ) der Sequenz gemäß Figur 1;
- (xviii) ss-DNA-Molekül mit einer Sequenz der Positionen 35422 bis 34205 (ORF 1) der Sequenz gemäß Figur 1;
- (xix) ss-DNA-Molekül mit einer Sequenz der Positionen 35522 bis 40147 (TubB) der Sequenz gemäß Figur 1;
- (xx) ss-DNA-Molekül mit einer Sequenz der Positionen 40144 bis 48021 (TubC) der Sequenz gemäß Figur 1;
- (xxi) ss-DNA-Molekül mit einer Sequenz der Positionen 48011 bis 58558 (TubD) der Sequenz gemäß Figur 1;
- (xxii) ss-DNA-Molekül mit einer Sequenz der Positionen 58551 bis 62096 (TubE) der Sequenz gemäß Figur 1;
- (xxiii) ss-DNA-Molekül mit einer Sequenz der Positionen 62103 bis 70616 (TubF) der Sequenz gemäß Figur 1;
- (xxiv) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x),

(xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii) oder (xxiii) unter stringenten Bedingungen hybridisierbar ist und insbesondere dieselbe Anzahl von Basen aufweist; und

(xxv) ss-DNA-Molekül, das mit einem ss-DNA-Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii) oder (xxiii) hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zu jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von diesem ss-DNA-Molekül hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und

(xxvi) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines Moleküls gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv) oder (xxv) komplementär ist.

Ferner betrifft die Erfindung ein ds-DNA-Molekül aus einem derartigen erfindungsgemäßen ss-DNA-Molekül und einem dazu komplementären Strang.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung ein ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:

- (i) ss-DNA-Molekül mit einer Sequenz der Positionen 35747 bis 36769 (Domäne C des tubB-Gens) der Sequenz gemäß Figur 1;
- (ii) ss-DNA-Molekül mit einer Sequenz der Positionen 37184 bis 39817 (Domäne A des tubB-Gens) der Sequenz gemäß Figur 1;
- (iii) ss-DNA-Molekül mit einer Sequenz der Positionen 38369 bis 39730 (Domäne NMT des tubB-Gens) der Sequenz gemäß Figur 1;
- (iv) ss-DNA-Molekül mit einer Sequenz der Positionen 39818 bis 40069 (Domäne PCP des tubB-Gens) der Sequenz gemäß Figur 1;
- (v) ss-DNA-Molekül mit einer Sequenz der Positionen 40372 bis 41397 (Domäne C des tubC-Gens) der Sequenz gemäß Figur 1;
- (vi) ss-DNA-Molekül mit einer Sequenz der Positionen 41824 bis 43215 (Domäne A des tubC-Gens) der Sequenz gemäß Figur 1;
- (vii) ss-DNA-Molekül mit einer Sequenz der Positionen 43216 bis 43461 (Domäne PCP des tubC-Gens) der Sequenz gemäß Figur 1;
- (viii) ss-DNA-Molekül mit einer Sequenz der Positionen 43552 bis 44574 (Domäne C des tubC-Gens) der Sequenz gemäß Figur 1;

- (ix) ss-DNA-Molekül mit einer Sequenz der Positionen 44980 bis 47631 (Domäne A des tubC-Gens) der Sequenz gemäß Figur 1;
- (x) ss-DNA-Molekül mit einer Sequenz der Positionen 46153 bis 47547 (Domäne NMT des tubC-Gens) der Sequenz gemäß Figur 1;
- (xi) ss-DNA-Molekül mit einer Sequenz der Positionen 47632 bis 47868 (Domäne PCP des tubC-Gens) der Sequenz gemäß Figur 1;
- (xii) ss-DNA-Molekül mit einer Sequenz der Positionen 48011 bis 49321 (Domäne KS des tubD-Gens) der Sequenz gemäß Figur 1;
- (xiii) ss-DNA-Molekül mit einer Sequenz der Positionen 49622 bis 50584 (Domäne AT des tubD-Gens) der Sequenz gemäß Figur 1;
- (xiv) ss-DNA-Molekül mit einer Sequenz der Positionen 51473 bis 52309 (Domäne KR des tubD-Gens) der Sequenz gemäß Figur 1;
- (xv) ss-DNA-Molekül mit einer Sequenz der Positionen 53066 bis 53980 (Domäne ER des tubD-Gens) der Sequenz gemäß Figur 1;
- (xvi) ss-DNA-Molekül mit einer Sequenz der Positionen 54158 bis 54460 (Domäne ACP des tubD-Gens) der Sequenz gemäß Figur 1;

- (xvii) ss-DNA-Molekül mit einer Sequenz der Positionen 54461 bis 55870 (Domäne HC des tubD-Gens) der Sequenz gemäß Figur 1;
- (xviii) ss-DNA-Molekül mit einer Sequenz der Positionen 56000 bis 57412 (Domäne A des tubD-Gens) der Sequenz gemäß Figur 1;
- (xix) ss-DNA-Molekül mit einer Sequenz der Positionen 57413 bis 57643 (Domäne PCP des tubD-Gens) der Sequenz gemäß Figur 1;
- (xx) ss-DNA-Molekül mit einer Sequenz der Positionen 58689 bis 59714 (Domäne C des tubE-Gens) der Sequenz gemäß Figur 1;
- (xxi) ss-DNA-Molekül mit einer Sequenz der Positionen 60156 bis 61697 (Domäne A des tubE-Gens) der Sequenz gemäß Figur 1;
- (xxii) ss-DNA-Molekül mit einer Sequenz der Positionen 61698 bis 61967 (Domäne PCP des tubE-Gens) der Sequenz gemäß Figur 1;
- (xxiii) ss-DNA-Molekül mit einer Sequenz der Positionen 62127 bis 63320 (Domäne KS des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxiv) ss-DNA-Molekül mit einer Sequenz der Positionen 63711 bis 64676 (Domäne AT des tubF-Gens) der Sequenz gemäß Figur 1;

- (xxv) ss-DNA-Molekül mit einer Sequenz der Positionen 64959 bis 65882 (Domäne KR des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxvi) ss-DNA-Molekül mit einer Sequenz der Positionen 65985 bis 67061 (Domäne CMT des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxvii) ss-DNA-Molekül mit einer Sequenz der Positionen 67242 bis 67829 (Domäne DH des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxviii) ss-DNA-Molekül mit einer Sequenz der Positionen 68247 bis 69128 (Domäne ER des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxix) ss-DNA-Molekül mit einer Sequenz der Positionen 69360 bis 69605 (Domäne PCP des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxx) ss-DNA-Molekül mit einer Sequenz der Positionen 69759 bis 70586 (Domäne TE des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxxi) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xiv), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix) oder (xxx) unter stringenten Bedingungen hybridisierbar ist und insbesondere dieselbe Anzahl von Basen aufweist;

- (xxxii) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix) oder (xxx) hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zur jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von diesem ss-DNA-Molekül hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und
- (xxxiii) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines Moleküls gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix), (xxx), (xxxi) oder (xxxii) komplementär ist.

Ferner betrifft die Erfindung ein ds-DNA Molekül aus einem derartigen ss-DNA-Molekül und einem dazu komplementärem Strang.

Ferner betrifft die Erfindung Varianten oder Mutanten, die aus einer Substitution, Insertion oder Deletion von Nucleotiden oder einer Inversion von Nucleotid-Segmenten eines erfindungsgemäßen ss-DNA-Moleküls oder eines erfindungsgemäßen ds-DNA-Moleküls resultieren, wobei diese Varianten und Mutanten Enzym-Varianten oder Enzym-Mutanten für die Produktion von Sekundärstoff(en) mit den eingangs geschilderten und für Tubulysine charakteristischen Eigenschaften codieren, insbesondere mit der cytostatischen Wirkung. Der Fachmann ist mit Massenscreening vertraut.

Ferner betrifft die Erfindung RNA

- (a) mit einer Sequenz entsprechend der eines erfindungsgemäßen ss-DNA-Moleküls oder
 - (b) mit einer Sequenz einer RNA gemäß (a), aber in Gegenrichtung (anti-sense), oder
 - (c) mit einer Sequenz einer RNA gemäß (a) oder (b) und mit einem dazu komplementären Strang,
- jeweils gegebenenfalls als Element eines rekombinanten Vektors.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung einen rekombinanten Vektor, insbesondere Expressionsvektor mit einem erfindungsgemäßen DNA-Molekül.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung eine Zelle, insbesondere zur Expression, in die ein erfindungsgemäßes DNA-Molekül oder ein erfindungsgemäßer Vektor integriert ist.

Die erfindungsgemäße Zelle kann sich von kultivierbaren Bakterien wie Myxobakterien wie *Angiococcus*, insbesondere *A. disciformis*, *Archangium*, insbesondere *A. gephyra*, *Escherichia coli*, *Pseudomonaden* oder *Actinomyceten* herleiten.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung eine Verwendung eines erfindungsgemäßen Vektors für die Transformation von Zellen oder Organismen zur transienten oder permanenten Expression eines oder mehrerer Proteine (Expressionsprodukt(e)), das (die) durch eine DNA (ssDNA oder dsDNA) des Vektors codiert wird (werden).

Gemäß einer weiteren Ausführungsform betrifft die Erfindung eine Verwendung einer erfindungsgemäßen Zelle zur enzymatischen

Biosynthese, Mutasynthese oder Partial-Synthese eines Tubulysins, insbesondere Tubulysin A, B, C, D, E und/oder F.

Gemäß einer weiteren Ausführungsform betrifft die Erfindung ein Expressionsprodukt eines erfindungsgemäßen DNA-Moleküls oder eines erfindungsgemäßen Vektors oder einer erfindungsgemäßen Zelle.

Die vorliegende Erfindung betrifft insbesondere ein Polynucleotid, das eine Sequenz wie in SEQ ID NO: 1, 18, 33 oder 36 definiert, oder ein Fragment davon enthält. SEQ ID NO: 1 und 18 beschreiben den (+)- bzw. den (-)-Strang des Tubulysin Biosynthese-Clusters von *Angiococcus disciformis*. SEQ ID NO: 33 ist eine Sequenz, die mehrere sich überlappende Gene des Clusters enthält. SEQ ID NO: 36 beschreibt eine Mutante von *Angiococcus disciformis*. Überraschenderweise wurde gefunden, dass diese Mutante ein Vielfaches der Tubulysin D-Produktion des Wildtyps zeigte. Die Tubulysin-Überexpression ist auf die Gesamtwirkung aller Tubulysin-Derivate bezogen sogar noch höher als die von Tubulysin D, was in keinsterweise zu erwarten war. Die Gene der SEQ ID NO: 36 sind offenbar an der negativen Regulation der Tubulysin Expression beteiligt. Diese Mutante ist auf Grund der erhöhten Expression aller Tubulysine besonders geeignet zur Produktion der erfindungsgemäßen Polypeptide. Antikörper gegen die Wildtyp-Expressionsprodukte dieser Sequenz können eingesetzt werden, um deren negativen Einfluss auf die Tubulysin Produktion auch in anderen Stämmen zu minimieren. Einen ähnlichen Effekt haben auch antisense-RNA bzw RNAi-Techniken, die mit der Wildtyp Sequenz der negativen Regulatorgene interagieren.

Die Fragmente des Polynucleotid können jede beliebige Teilsequenz und Länge haben, bevorzugt sind jedoch solche Fragmente, die Proteine codieren. Die Polynucleotide der vorliegenden Erfindung schließen ebenfalls ein, sind jedoch nicht beschränkt auf, ein Polynucleotid, das an den komplementären Strang der offenbaren Nucleotidsequenzen unter moderat stringenten oder stringenten Bedingungen hybridisiert; ein Polynucleotid, das eine Allel-Variante irgendeines oben beschriebenen Polynucleotids ist; ein Polynucleotid, das ein Spezies-Homolog irgendwelcher der hier offenbaren Proteine codiert; oder ein Polynucleotid, das ein Polypeptid codiert, das eine zusätzliche spezifische Domäne oder eine Trunkierung bzw. Verkürzung der offenbaren Proteine aufweist.

Der Begriff "CDS" bezeichnet eine Sequenz von Nucleotiden, die der Sequenz von Aminosäuren in einem Protein entspricht, also die Aminosäuren-codierenden Sequenzbereiche, inklusive des jeweiligen Start- und Stoppcodons.

In einer bevorzugten Ausführungsform ist das erfindungsgemäße Polynucleotid ein Fragment, das eine im Sequenzprotokoll definierte CDS ist.

Die vorliegende Erfindung betrifft weiterhin einen Vektor, der ein Polynucleotid wie oben beschrieben enthält. Vektoren für verschiedene Zwecke sind im Stand der Technik bekannt, ebenso die Techniken zur Subklonierung von Polynucleotiden in solche Vektoren. Diese werden beispielsweise in der neuen Ausgabe von Molecular Cloning: A Laboratory Manual, (Sambrook et al., (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Gene Transfer Vectors for Mammalian Cells

(Miller & Calos, eds.); Current Protocols in Molecular Biology and Short Protocols in Molecular Biology, 3rd Edition (F. M. Ausubel et al., eds.); Recombinant DNA Methodology (R. Wu ed., Academic Press) oder "A Practical Guide To Molecular Cloning" beschrieben. Beispiele für Vektoren finden sich unter anderen in Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory).

Bei dem Vektor handelt es sich vorzugsweise um einen Expressionsvektor, also im Allgemeinen ein Plasmid, einen Phagen, ein Virus oder einen Vektor zum Exprimieren eines Polypeptids aus einer DNA (RNA) Sequenz. Ein Expressionsvektor kann eine Transkriptionseinheit umfassen, die eine Anordnung des Folgenden aufweist: (1) ein genetisches Element oder Elemente mit einer regulatorischen Rolle in der Genexpression, beispielsweise Promotoren oder Enhancer, (2) eine Struktursequenz oder codierende Sequenz, die in mRNA transkribiert und in ein Protein translatiert wird und (3) geeignete Transkriptionsstart- und -terminationssequenzen. Struktureinheiten, die zur Verwendung in Hefen oder eukaryontischen Expressionssystemen vorgesehen sind, schließen vorzugsweise eine Leadersequenz ein, die die extrazelluläre Sekretion eines translatierten Proteins durch einen Wirt ermöglicht. Es kann alternativ, wenn ein rekombinantes Protein ohne eine Leader- oder Transportsequenz exprimiert wird, einen N-terminalen Methionin-Rest einschließen. Dieser Rest kann, muss aber nicht anschließend von dem exprimierten rekombinanten Protein abgespalten werden, um das Endprodukt zu erhalten.

Des Weiteren betrifft die vorliegende Erfindung eine Zelle, die einen solchen Vektor enthält. Der Vektor kann über die bekannten Techniken wie beispielsweise Transfektion, Elektroporation,

Lipofektion usw. in die Zelle eingebracht werden. Bei viralen Vektoren ist auch eine Infektion möglich. Bei den Zellen kann es sich um eukaryotische als auch prokaryotische Zellen handeln. Auch die Methoden zur Selektion und Propagierung der Zellen, die den Vektor enthalten sind dem Fachmann bekannt. Beispiele für die Kultivierung von Zellen tierischen Ursprungs sind u.a. in Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987) zu finden.

Eine weitere Ausführungsform betrifft ein Polypeptid, das wenigstens eine Sequenz wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder ein Fragment und/oder Derivat davon enthält. Das Polypeptid kann sowohl durch Expression eines Polynucleotids als auch durch chemische Synthese bereitgestellt werden.

Die Aminosäuresequenzen der vorliegenden Erfindung umfassen ebenfalls alle Sequenzen, die sich von den hier offenbarten Sequenzen durch Aminosäureinsertionen, -deletionen und -substitutionen unterscheiden.

Aminosäure- „Substitutionen“ sind vorzugsweise das Ergebnis des Ersetzens einer Aminosäure durch eine andere Aminosäure mit ähnlichen strukturellen und/oder chemischen Eigenschaften, das heißt konservative Aminosäure-Austausche. Aminosäure-Substitutionen können auf der Grundlage einer Ähnlichkeit in der Polarität, Ladung, Löslichkeit, Hydrophobie, Hydrophilie und/oder der amphipatischen Natur der einbezogenen Reste vorgenommen werden. Beispielsweise schließen unpolare (hydrophobe) Aminosäuren Alanin, Leucin, Isoleucin, Valin, Prolin, Phenylalanin, Tryptophan und Methionin ein; polare neutrale Aminosäuren schließen Glycin, Serin, Threonin, Cystein,

Tyrosin, Asparagin und Glutamin ein; positiv geladene (basische) Aminosäuren schließen Arginin, Lysin und Histidin ein; und negativ geladene (saure) Aminosäuren schließen Asparaginsäure und Glutaminsäure ein.

"Insertionen" oder "Deletionen" bewegen sich typischerweise im Bereich von 1-3 Aminosäuren. Die erlaubte Variation kann experimentell bestimmt werden, indem systematisch Insertionen, Deletionen oder Substitutionen von Aminosäuren in einem Polypeptidmolekül unter Verwendung von DNA-Rekombinationstechniken vorgenommen und die sich ergebenden rekombinanten Varianten bezüglich ihrer Aktivität untersucht werden. Dazu ist für den Fachmann nicht mehr als die Durchführung von Routineexperimenten erforderlich.

Beispielsweise kann das Polypeptid auch in Form eines chimären Polypeptids, codiert durch ein Fusionsgens, vorliegen, dass wenigstens eine weitere Sequenz enthält. Diese zusätzliche Sequenz kann z.B. dazu dienen, die Aufreinigung des Expressionsprodukts zu erleichtern oder dem Expressionsprodukt eine zusätzlich Funktion zu verleihen.

Beispiele für zusätzliche Sequenzen, die die Aufreinigung erleichtern sind sogenannte Tags, die dem Fachmann bekannt sind, wie z.B. das his-tag.

Die vorliegende Erfindung betrifft außerdem die Verwendung wenigstens eines Polynucleotids wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens eines Fragments davon und/oder wenigstens eines Polypeptids wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens eines Fragments davon zur Herstellung einer pharmazeutischen Zusammensetzung zur Behandlung von

unerwünschtem Zellwachstums oder unerwünschter Zellvermehrung in einem Individuum. Die Zusammensetzung kann beispielsweise einen geeigneten Vektor enthalten, zusammen mit Hilfsfaktoren, die die Expression eines Tubulysins, vorzugsweise in den unerwünschten Zellen, ermöglichen und dadurch ein weiteres Wachstum bzw. eine weitere Vermehrung dieser Zellen verhindern. Die Zusammensetzung kann auch erfindungsgemäße Zellen enthalten, die mit einem Vektor, beispielsweise einen Tubulysin-exprimierenden Vektor, transfiziert wurden.

In einer bevorzugten Ausführungsform ist das unerwünschte Zellwachstum oder die unerwünschte Zellvermehrung ein Tumor. Der Tumor kann sowohl eine gutartige als auch eine bösartige Geschwulst sein.

In einer weiteren Ausführungsform ist das unerwünschte Zellwachstum eine pathogene Infektion. Das Pathogen kann hierbei sowohl einzellig als auch mehrzellig sein. Dies schließt auch Infektionen mit Pilzen, wie beispielsweise Candida oder Aspergillus, und Infektionen mit Parasiten, wie beispielsweise Trypanosomen oder Schistosomen ein. In einer bevorzugten Ausführungsform der Verwendung ist die pathogene Infektion eine Mykose, Malaria oder eine parasitäre Erkrankung.

Die Erfindung betrifft ferner eine pharmazeutische Zusammensetzung, enthaltend wenigstens ein Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens ein Fragment davon und/oder wenigstens ein Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens ein Fragment davon. Die Zusammensetzungen enthalten eine therapeutisch wirksame Menge bzw. Dosis des jeweiligen Wirk- bzw. Inhaltsstoffes. Eine therapeutisch

wirksame Dosis betrifft diejenige Menge der Verbindung, die ausreicht, um eine Linderung der Symptome, beispielsweise eine Behandlung, Heilung, Prävention oder Linderung derartiger Zustände, insbesondere die Hemmung oder Verhinderung von unerwünschtem Zellwachstum und Zellvermehrung in einem Patienten zu ergeben. Geeignete Verabreichungswege schließen beispielsweise eine parenterale Verabreichung, einschließlich intramuskulärer und subkutaner Injektionen, ebenso wie intrathekaler, direkt intraventrikulärer, intravenöser, intraperitonealer Injektionen ein.

In einer weiteren Ausführungsform enthält diese pharmazeutische Zusammensetzung wenigstens einen pharmazeutisch verträglichen Trägerstoff. Eine solche Zusammensetzung kann weiterhin (zusätzlich zum Inhaltsstoff und dem Träger) Verdünnungsmittel, Füllmittel, Salze, Puffer, Stabilisatoren, Lösungsvermittler und andere Materialien enthalten, die im Stand der Technik wohl bekannt sind. Der Begriff "pharmazeutisch verträglich" bedeutet ein nicht-toxisches Material, das die Wirksamkeit der biologischen Aktivität des aktiven Inhaltsstoff (der aktiven Inhaltsstoffe) nicht stört. Die Eigenschaften des Trägers hängen vom Verabreichungsweg ab. Die therapeutische Zusammensetzung kann weiterhin weitere Mittel bzw. Wirkstoffe enthalten, die die Aktivität bzw. Wirksamkeit oder Anwendung bei Behandlung verbessern bzw. erleichtern. Solche zusätzlichen Faktoren und/oder Mittel können in der therapeutischen Zusammensetzung enthalten sein, um eine synergistische Wirkung zu erzeugen oder um Nebenwirkungen zu minimieren.

Techniken zur Formulierung, Zubereitung und Verabreichung der Verbindungen der vorliegenden Erfindung sind in "Remington's

Pharmaceutical Sciences", Mack Publishing Co., Easton, PA, neueste Ausgabe, zu finden.

Die vorliegende Erfindung betrifft außerdem ein Verfahren zur Herstellung von Tubulysinen und Tubulysin-Biosynthese Proteinen, umfassend die Schritte:

- (a) exprimieren von wenigstens einem Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens einem Fragment davon und/oder wenigstens einem Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens einem Fragment davon, und
- (b) aufreinigen der Expressionsprodukte.

Methoden zur Expression von Proteinen sind dem Fachmann bekannt und können der einschlägigen Literatur entnommen werden, beispielweise aus Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.) oder Recombinant DNA Methodology (R. Wu ed., Academic Press). Für das Aufreinigen von Expressionsprodukten sind dem Fachmann eine Vielzahl von Methoden bekannt. Neben chromatographischen Methoden wie beispielsweise Affinitätschromatographie oder HPLC können auch immunologische Verfahren wie beispielsweise immobilisierte Antikörper gegen ein Epitop auf dem Expressionsprodukt, z.B. einem His-tag, zum Aufreinigen der Produkte verwendet werden.

In einer bevorzugten Ausführungsform erfolgt die Expression in prokaryotischen oder eukaryotischen Zellen und/oder durch in vitro Expression. Die Expression von Polypeptiden in prokaryotischen oder eukaryotischen Zellen ist ein häufig angewandtes Verfahren und wird im Allgemeinen mittels eines Expressionsvektors, wie oben beschrieben, erreicht. Für die in vitro Expression sind ebenfalls bereits Vektoren beschrieben.

Diese und die notwendigen Faktoren sind kommerziell in Form von Kits erhältlich, beispielsweise von BioRad, Stratagene, Invitrogen und Clontech.

Desweiteren stellt die Erfindung ein Verfahren zum Auffinden von Genen, die an der Biosynthese von Tubulysinen beteiligt sind, bereit. Das Verfahren umfasst die folgenden Schritte:

- (a) hybridisieren von wenigstens einem Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens einem Fragment davon mit DNA, RNA, und/oder cDNA einer Spezies, die nicht mit *Angiococcus disciformis* identisch ist, und
- (b) isolieren und charakterisieren der hybridisierten DNA, RNA, und/oder cDNA.

Die Hybridisierung kann unter unterschiedlich stringenten Bedingungen durchgeführt werden.

Die Stringenz der Hybridisierung, wie hierin verwendet, betrifft Bedingungen, unter denen Polynucleotid-Doppelstränge stabil sind. Wie dem Fachmann bekannt ist, ist die Stabilität eines Doppelstranges eine Funktion der Natriumionenkonzentration und der Temperatur (siehe beispielsweise Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed. (Cold Spring Harbor Laboratory, (1989)). Die Stringenzniveaus, die zur Hybridisierung verwendet werden, können vom Fachmann leicht abgewandelt werden.

Der Begriff „schwach stringente Hybridisierung“ bezeichnet Bedingungen, die einer Hybridisierung in 10% Formamid, 5 x Denharts Lösung, 6 x SSPE, 0,2 % SDS bei 42°C, gefolgt von Waschen in 1 x SSPE, 0,2 % SDS bei 50°C äquivalent sind.

Denhart's Lösung und SSPE sind dem Fachmann genauso wie andere geeignete Hybridisierungspuffer wohl bekannt.

Eine „moderat stringente Hybridisierung“ bedeutet Bedingungen, die es der DNA erlauben, an eine komplementäre Nucleinsäure zu binden, die ungefähr 60 % Identität, vorzugsweise ungefähr 75 % Identität, besonders bevorzugt ungefähr 85 % Identität zu dieser DNA aufweist; wobei eine Identität von mehr als ungefähr 90 % zu dieser DNA besonders bevorzugt wird. Moderat stringente Bedingungen sind vorzugsweise Bedingungen, die eine Hybridisierung in 50 % Formamid, 5 x Denharts Lösung, 5 x SSPE, 0,2 % SDS bei 42°C gefolgt von Waschen in 0,2 x SSPE, 0,2 % SDS bei 65°C äquivalent sind.

Hochstringente Hybridisierung bedeutet Bedingungen, die die Hybridisierung nur von solchen Nucleinsäuresequenzen ermöglichen, die in 0,018 M NaCl bei 65°C stabile Doppelstränge bilden (d.h., wenn ein Doppelstrang in 0,018 M NaCl bei 65°C nicht stabil ist, ist er unter den hier beschriebenen/definierten hochstringenten Bedingungen nicht stabil).

Weiterhin können Nucleinsäure-Hybridisierungstechniken verwendet werden, um eine Nucleinsäure zu identifizieren und zu gewinnen, die von der vorliegenden Erfindung umfasst ist. Kurz gesagt kann jede Nucleinsäure mit einer gewissen Homologie zu einer in dieser Erfindung offenbarten Sequenz oder einem Fragment davon, als Sonde zur Identifizierung einer ähnlichen Nucleinsäure durch Hybridisierung unter moderat stringenten bis hochstringenten Bedingungen verwendet werden. Solche ähnlichen Nucleinsäuren können dann isoliert, sequenziert und analysiert werden, um zu bestimmen, ob sie von der vorliegenden Erfindung umfasst werden.

Die vorliegende Erfindung stellt außerdem einen Kit zur Herstellung von Tubulysinen, enthaltend:

- (a) wenigstens ein Polynucleotid, enthaltend eine Sequenz wie in SEQ ID NO: 1, 18, 33 oder 36 definiert oder ein Fragment davon und/oder wenigstens einen Vektor der ein solches Polynucleotid enthält oder
- (b) geeignete Medien und Puffer zur Vermehrung von Zellen, die eine Expression des Polynucleotids und/oder Vektors erlauben, und
- (c) geeignete Mittel zum Aufreinigen des/der Expressionsprodukte(s)

bereit.

Auf Grund ihrer Wirkung auf das Tubulin-Skelett und ihrer Cytotoxizität, insbesondere bei Pilzen, eignen sich Tubuline auch als Desinfektionsmittel, das eine Kontamination mit Tubulin-enthaltenden Zellen abbauen bzw. verhindern kann. Die Erfindung betrifft deshalb auch die Verwendung einer Zusammensetzung enthaltend wenigstens ein Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens ein biologisch aktives Fragment oder Derivat davon als Desinfektionsmittel. Neben dem oben definierten Polypeptid können im Desinfektionsmittel auch andere desinfizierend wirkende Stoffe enthalten sein, solange sie die Wirkung des erfindungsgemäßen Polypeptids nicht inhibieren. Außerdem kann das Desinfektionsmittel weitere Hilfsstoffe wie beispielsweise Puffer, Wasser, Farbstoffe, Duftstoffe, Stabilisatoren, Trägerstoffe usw. enthalten. In einer bevorzugten Ausführungsform ist die Zusammensetzung flüssig oder pulverförmig.

Entsprechend betrifft die Erfindung auch Desinfektionsmittel wie oben definiert.

Soweit keine anderen Definitionen angegeben sind, weisen alle technischen und wissenschaftlichen Begriffe, die hierin verwendet werden, dieselbe Bedeutung auf, wie sie üblicherweise vom Fachmann auf dem Gebiet, an den sich diese Erfindung wendet, verstanden werden. Alle Veröffentlichungen, Patentanmeldungen, Patente und anderen hierin erwähnten Referenzen sind durch Bezugnahme in ihrer Gesamtheit mit aufgenommen. Im Falle eines Konfliktes wird jedoch die vorliegende Beschreibung einschließlich der Definitionen entscheiden. Zusätzlich sind die Materialien, Methoden und Beispiele lediglich veranschaulichend und sollen nicht als einschränkend aufgefasst werden.

1. Identifizierung des Tubulysin-Biosyntheseclusters in *Angiococcus disciformis* An d48 durch mariner Transposon-Mutagenese mittels pMycoMar

Die Identifizierung des Tubulysin-Biosyntheseclusters wurde durch die Erstellung einer Transposon-Mutantenbank aus *Angiococcus disciformis* An d48 mittels pMycoMar durchgeführt.

Rubin & Mekalanos (Proc. Natl. Acad. Sci. USA, 96 (1999), 1645 - 1650) entwickelten aus dem mariner Element Himar1 das Plasmid pMycoMar, welches ein einfaches Transpositionssystem darstellt, das effizient Bakterien *in vivo* infizieren und Insertions-Mutanten generieren kann. Dieses Plasmid enthält das Mini-Transposon *magellan4*, bei dem das Tn5 Kanamycin-Resistenzgen und der oriR6K von den inverted repeats des Himar1 flankiert sind. Zusätzlich ist die Himar1 Transposase unter der transkriptionalen Kontrolle des T6 Promotors in das mycobakterielle Temperatur-sensitive Replicon pPR23 kloniert worden. Das pMycoMar codiert ebenfalls ein Gentamycin-Resistenzgen.

Das Himar1 zeichnet sich bei der Transposition durch eine TA Dinucleotid-Erkennungssequenz aus. Daher kann es zufällig in ein Wirts-Genom integrieren und statistisch gesehen alle aktiven Gene durch eine Insertions-Mutation ausschalten. Aufgrund dieser Tatsache sollte eine Mutantenbank aus An d48 generiert und das Tubulysin-Biosynthesecluster durch eine Knockout-Mutante identifiziert werden.

Alternativ kann man auch von *Archangium gephyra* DSM 11092 ausgehen und nach einem Protokoll von Biozym Diagnostic (Oldendorf, DE; Katalog TSM99K2; pEZ::TN<KAN-2> Tnp Transposome-Kit) arbeiten.

1.1 Generierung der Mutantenbank

Für eine Elektrotransformation von *A. disciformis* An d48 wurden zwei verschiedene Protokolle verwendet. Diese Protokolle wurden für die Myxobakterien *Stigmatella aurantiaca* (Stamm & Plaga, Arch. Microbiol., 172 (1995), 483 - 494) und *Myxococcus xanthus* (Kashefi & Hartzell, Mol. Microbiology, 15 (1995) 483 - 494) etabliert. Die beiden Methoden zeigten in der Transformationseffizienz des *A. disciformis* An d48 keinen Unterschied, so dass die Elektrotransformation zur Erstellung der Transposonbank nach dem Protokoll für *Stigmatella aurantiaca* durchgeführt wurde. Im folgenden sind die beiden Protokolle aufgeführt.

1.1.1. Elektrotransformation von *Angiococcus disciformis* An d48 nach *Stigmatella aurantiaca*-Protokoll

Eine in 50 ml Tryptonmedium (10 g Trypton; 2 g $MgSO_4$; 0,1 % Vitamin B12 [10 ng/ml]; 0,2 % Glucose auf 1 l Medium; pH 7,2) gewachsene *A. disciformis* Kultur wird bis 2×10^8 Zellen / ml bei 30°C kultiviert. Ausgehend von einer Generationszeit von 6 Stunden, wurde diese Kultur am Vortag so angeimpft, dass rechnerisch diese Zelldichte erreicht wird. Die Kultur wird bei 20°C abzentrifugiert (20min; 4000rpm) und die Zellen im gleichen Volumen Waschpuffer (5 mM HEPES/NaOH, 0,5 mM $CaCl_2$, pH 7,2) resuspendiert. Nach erneuter Zentrifugation wird in 25 ml Puffer resuspendiert und erneut abzentrifugiert. Vor diesem Zentrifugationsschritt wird die absolute Zellzahl innerhalb der 25 ml bestimmt, so dass rechnerisch 1×10^9 Zellen/ 40 μ l resuspendiert werden.

Elektroporationsbedingungen:

1-3 µg DNA und 40 µl Zellsuspension werden gemischt und in eine auf Eis gekühlte Elektroporationskuvette (0,1 cm) gegeben. Die Elektroporation wird bei 200 Ω, 25 mF und 0,85 kV / cm durchgeführt.

Direkt nach der Elektroporation wird 1 ml Tryptonmedium zugegeben. Nach Transfer in 50 ml Tryptonmedium werden die Zellen für eine phänotypische Expression 6 h bei 30°C geschüttelt. Danach wird die Kultur abzentrifugiert (20 min, 4000 rpm, 20°C) und in 1 ml Trypton resuspendiert. Ausgehend von einer 100 %-igen Überlebensrate der Zellen wird eine Verdünnungsreihe erstellt und $1 \cdot 10^8$ - $1 \cdot 10^4$ Zellen werden mit 3 ml Trypton-Softagar auf Kanamycin haltigen (50 µg / ml) Tryptonplatten plattiert. Diese Platten werden bei 30°C inkubiert und nach 5 - 8 Tagen sind die ersten Klone zu sehen.

1.1.2. Elektrotransformation von *Angiococcus disciformis* An d48 nach *Myxococcus xanthus*-Protokoll

Die Wachstumsbedingungen der Vor- und Hauptkultur sowie Zentrifugationen und das abschließende Einengen der Zellzahl wurden genauso durchgeführt wie unter 1.1.1. angegeben ist. Dies ist abweichend vom Standardprotokoll für *Myxococcus xanthus* optimiert worden.

Elektroporationsbedingungen:

1-3 µg DNA und 40 µl Zellsuspension werden gemischt und in eine auf Eis gekühlte Elektroporationskuvette (0,1 cm) gegeben. Die Elektroporation wird bei 400 Ω, 25 µF und 0,65kV / cm durchgeführt.

Direkt nach der Elektroporation wird 1 ml Tryptonmedium zugegeben und in einem 1,5 ml Eppendorf-Reagenzgefäß für 6 h bei 30°C geschüttelt. Ausgehend von einer 100%-igen Überlebensrate der Zellen wird eine Verdünnungsreihe erstellt und $1 \cdot 10^8$ - $1 \cdot 10^4$ Zellen mit 3 ml Trypton-Softagar auf Kanamycin haltigen (50 µg / ml) Tryptonplatten plattiert. Diese Platten werden bei 30°C inkubiert und nach 5 - 8 Tagen sind die ersten Klone / Mutanten zu sehen, die mittels Impföse gepickt wurden.

1.2 Kultivierung der Transposonmutanten

Die Mutanten wurden in 96'er Mikrotiter-Platten in 200 µl M7-Medium (5 g Probion; 1 g CaCl_2 ; 1 g MgSO_4 ; 1g Hefeextrakt; 5 g Stärke; 10 g HEPES; 0,1 % Vitamin B12 [10 ng / ml] auf 1 l Medium; pH7,4) bei 32°C inkubiert und nach 10 Tagen eine Kopie der gesamten Bank erstellt. Dafür wurden 50 µl Kultur jeder Mutante in neue Mikrotiter-Platten mit 100 µl M7-Medium transferiert. Nach weiteren sieben Tagen Inkubation wurde eine Kopie bei -80°C als Dauerkulturen eingefroren. Die verbleibende Kopie der Bank wurde extrahiert, der Extrakt mittels Toxizitätstest auf generierte Tubulysin Knockout-Mutanten untersucht.

Bei der Identifizierung von Mutanten, die in dieser Analytik Veränderungen zum Wildtyp zeigten (keine Zellkernfragmentierung), wurden diese aus der Dauerkultur rekultiviert. Zur Kontrolle der erzielten Ergebnisse sollten 50 ml M7-Medium Großkulturen der entsprechenden Mutanten erneut getestet werden. Bei eventuellen Tubulysin Knockout-Mutanten wurden die Extrakte zunächst über einen HPLC-Lauf fraktioniert und anschließend die Fraktionen mittels Toxizitätstest auf Tubulysin untersucht. Durch die vorhergehende Fraktionierung wird eine Maskierung der Tubulysin-Wirkung durch Myxothiazol vermieden. Da die beiden

Sekundärmetabolite verschiedene Retentionszeiten bei der Elution von einer C-14 Säule besitzen, liegen beide im folgenden Toxizitätstest in unterschiedlichen Fraktionen vor.

1.3. Toxizitätstest

Die Mini-Kulturen aus den 96 Mikrotiterplatten wurden nach Kultivierung durch Stickstoff-Begasung auf einem Heizblock bei 37°C eingetrocknet. Danach wurden die Zellpellets über 2 h in 100 µl Methanol resuspendiert und jeweils 10 µl für den folgenden Toxizitätstest eingesetzt, um eine Tubulysin-Produktion der jeweiligen Mutante detektieren zu können.

Für diesen Test werden L929 Zellen in DMEM-Medium (Invitrogen, Groningen) bei 37°C kultiviert und danach mittels Zellschaber vorsichtig geerntet. Diese Zellsuspension wird anschließend 1 : 10 mit DMEM verdünnt und 120 µl pro Öffnung einer 96 Mikrotiterplatte verteilt. Diese werden anschließend mit den 10 µl Zellextrakt der einzelnen Transposon-Mutanten versetzt und für fünf Tage bei 37°C inkubiert. Nach dieser Inkubationszeit werden die L929 Zellen mikroskopisch auf Zellkernfraktionierung untersucht, was ein Zeichen für Tubulysineinwirkung ist. Bei Zellen, die keine Zellkernfraktionierung zeigten, wurden die entsprechenden Mutanten als mutmaßliche Tubulysin Knockout-Mutanten identifiziert. Die Extrakte dieser Mutanten wurden in 50 ml M7-Medium (+ 1ml Absorber-Harz XAD-16 von Rohm & Haas) angezogen und die Zellkerne der L929-Zellen nach durchgeführtem Toxizitätstest zusätzlich durch eine Anfärbung des Chromosoms mittels DAPI-Färbung auf eine Zellkernfraktionierung bzw. Tubulysin-Produktion untersucht.

1.4 Bestimmung des Integrationsgenortes der Tubulysin Knockout-Mutanten in der An d48 mariner vermittelten Mutantenbank mittels Transposon-recovery

In der generierten Mutantenbank konnten mittels Toxizitätstest fünf Mutanten identifiziert werden (MutT176, 524, 781, 794 und 929), die kein Tubulysin produzierten. Dieses Ergebnis konnte nach Rekultivierung der Mutanten aus der Dauerkultur und erneuter Analyse bestätigt werden. Um Informationen zu erhalten, in welchen Bereich des Genoms das Himarl Element transponiert ist, wurde ein Transposon-recovery durchgeführt. Bei dieser Methodik wird die chromosomale DNA der jeweiligen Mutante mit verschiedenen Restriktionsenzymen geschnitten, die nicht innerhalb der bekannten *magellan4* Sequenz schneiden. Die restringierte DNA wird ligiert, nach Transformation in DH5 α / λ pir-Zellen wird auf Kanamycin haltigen LB-Platten bei 37°C inkubiert. Auf diesen Platten können nur solche *E.coli* Zellen wachsen, die ein Plasmid mit dem *magellan4* und somit dem Tn5 Kanamycin-Resistenzgen enthalten. Solch ein Plasmid enthält an den Enden des Transposons chromosomale DNA aus An d48. Diese Plasmide können in den *E.coli* Zellen DH5 α / λ pir replizieren, da innerhalb der Transposon Sequenz der *oriR6K* sitzt. Das Transposon wurde so aus dem jeweiligen Genom isoliert und mit den Primern K388 und K389 ansequenziert. Die erhaltenen Sequenzen wurden dann gegen die Genbank auf Homologien zu bekannten Genen hin untersucht und zeigten dabei hohe Ähnlichkeiten zu nicht ribosomalen Peptidsynthetasen (NRPS) aus bekannten Sekundärmetabolit-Biosynthesegenclustern wie denen des Myxothiazols, Nostopeptolids und Saframycins. Diese Analysen waren eindeutige Hinweise dafür, dass es sich dabei um Sequenzfragmente aus dem gesuchten Tubulysin-Gencluster handelt. Durch Restriktions- und Southernanalysen wurde die Größe der

einzelnen Transposon-Plasmide und ihre relativen Integrationsorte zueinander (innerhalb des Genclusters) bestimmt.

1.4.1. Transposon-recovery

Isolieren chromosomaler DNA nach Standardprotokollen aus 50 ml Tryptonmedium-Kultur jeder *A. disciformis* An d48 Mutante. Von dieser DNA werden 5 µg für die folgende Ausklonierung des Transposons verwendet, indem zunächst eine Restriktion durchgeführt wird. Dabei wurden die Enzyme *NotI* und *BamHI* verwendet, die keine Restriktionsstelle innerhalb des *magellan4* haben und statistisch relativ häufig in GC-reicher DNA schneiden sollten.

Verdau der genomischen DNA mit *NotI* bzw. *BamHI*:

- 5 µg DNA
- + 3 µl 10 x NEB Puffer
- + 3 µl 100 x BSA
- + 10 U Restriktionsenzym (*BamHI* bzw. *NotI*)
- + x µl dest. H₂O

30 µl Ansatz für 3 h bei 37°C inkubieren ⇒ wiederum 10 U Enzym zu den Restriktionsansätzen zugeben und weitere 2 h bei 37°C inkubieren.

Fällen der restringierten DNA und folgende Ligation

Der gesamte Restriktionsansatz von 30 µl wird mit 1 Vol. Chloroform/Phenol versetzt und für 10 min. zentrifugiert (13.200 rpm; 20°C). Der Überstand wird in ein neues Reaktionsgefäß überführt und mit 1/10 Vol. 3 M NaOAc und 2,5 Vol. 100% EtOH versetzt. Zum fällen der DNA wird das

Reaktionsgefäß für 1 h bei -20°C inkubiert und anschließend für 30 min. zentrifugiert (13.200 rpm; 4°C). Der Überstand wird verworfen und das Pellet dreimal mit 70% EtOH gewaschen, wobei man jeweils 5 min. zentrifugiert (13.200 rpm; 20°C). Nach dem Verwerfen des Überstandes wird das Pellet bei 37°C getrocknet und in 15 µl H₂O resuspendiert. Für die anschließende Ligation werden die gesamten 15 µl der gefällten DNA verwendet.

Ligationsansatz:

	15 µl DNA
+	4 µl 5 x Ligasepuffer (NEB)
+	<u>1 µl NEB Ligase</u>

20 µl Ansatz über Nacht bei 16°C inkubiert

⇒ wiederum 1 µl Ligase zu den Ligationsansätzen und UN
bei 16°C inkubieren.

Elektrotransformation der Ligationsansätze in den *E. coli* Stamm DH5 α - λ pir

1-3 μ l der Ligationsansätze und 50 μ l DH5 α - λ pir Zellen werden gemischt und in eine auf Eis gekühlte Elektroporationsküvette (0,1 cm) gegeben. Die Elektroporation wird bei 200 Ω , 25 mF und 1,25kV / cm durchgeführt. Die Zellen werden danach in 1 ml LB-Medium (10 g Trypton; 10 g NaCl; 5 g Hefeextrakt auf 1 l Medium) aufgenommen und für 1 h bei 37°C inkubiert. Danach werden sie auf Kanamycin-haltigen (50 μ g / ml) LB-Platten ausplattiert. Nach einem Tag Inkubation bei 37°C können die Klone gepickt werden. Dabei können nur Zellen anwachsen, die ein Transposon-Plasmid beinhalten und damit eine Tn5-Kan^R vermittelte Resistenz besitzen.

1.5 Sequenzauswertung des Tubulysin Biosynthese-Genclusters aus pMutT794/NotI

Das Transposon-Plasmid pMutT794/NotI enthält 52985 bp chromosomale DNA aus *Angiococcus disciformis* An d48. Zusammen mit dem HimarI Minitransposon *magellan4* (2199 bp), das bei Basenpaar 37317 bp in das Plasmid integriert ist, wurden 55184 bp sequenziert. Insgesamt 21760 bp stammen aus codierenden Genen des Tubulysin-Genclusters und sowie 31219 bp weitere codierende Gene. Bei diesen ORF's handelt es sich zum Teil um Regulatorgene, die die Expression des Tubulysins beeinflussen können. Sequenzvergleiche mit den Transposon-Plasmiden der anderen Tubulysin Knockout-Mutanten zeigten, dass das *magellan4* bei den Mutanten MutT781 (36975bp) und MutT929 (36197 bp) innerhalb von 1658 bp zu MutT794 in das Biosynthese-Gencluster transponiert sind.

Auf der Sequenz ist der Start des Tubulysin-Genclusters mit drei NRPS-Modulen (*tubA-C*), einem Cyclodeaminase codierendem Gen (*tubZ*) und einem PKS-Modul (*tubD*) enthalten. Desweiteren liegen innerhalb des Genclusters ein Anionentransporter codierendes Gen (ORF1), das zum Austransport des Tubulysins aus der Zelle dient und ein weiterer ORF (ORF2). Die grundsätzliche Anordnung der Gene, sowie der einzelnen Domänen mit einer N-Methyltransferase innerhalb der Adenylierungsdomänen (A) von *tubB* und *tubC*, entsprechen dem typischen Aufbau des Genclusters und der damit verbundenen Tubulysin-Biosynthese. Die Methyltransferase-Domänen (NMT) sind aber im Gegensatz zu bekannten Genclustern-Strukturen nicht zwischen der Adenylierungs- und Thiolierungs-Domäne (PCP) lokalisiert, sondern zwischen A8 und A9 innerhalb der Adenylierungs-Domäne (A) (hoch konservierte Regionen innerhalb der Adenylierungs-Domänen von NRPS; Konz & Marahiel, Chem. Biology, 6 (1999) R39 - R48). *TubA*, codiert eine unvollständige

Kondensationsdomäne, die für die Biosynthese theoretisch nicht benötigt wird. Die am Ende der bekannten Sequenz liegende Polyketidsynthase (PKS) enthält eine Ketoacylsynthase- (KS), Acyltransferase- (AT) und Ketoreduktase- (KR) Domäne.

Die restliche Sequenz des Tubulysin-Biosynthesegenclusters wurde aus einer Cosmidbank von An d48 (hinterlegt bei der DSMZ) unter Standardbedingungen identifiziert. Das auf der ersten Hälfte der Sequenz endende PKS-Modul (*tubD*) wird fortgesetzt von den schon erwähnten KS, AT und KR-Domänen und beinhaltet weiterhin eine Enoylreduktase (ER) und ein Acyl Carrier Protein (ACP). In der folgenden Sequenz von *tubD* wird eine NRPS codiert, die eine Heterozyklisierungs- (HC), Adenylierungs- (A) und Peptidyl Carrier Protein- (PCP) Domäne trägt. Des weiteren folgen die Gene *tubE* und *tubF*. Das Gen *tubE* codiert eine NRPS mit den Domänen C, A und PCP. Auf *tubF* wird eine PKS mit der folgenden Domänen-Anordnung codiert: Ketoacylsynthase (KS), Acyltransferase (AT), Ketoreduktase (KR), C-Methyltransferase (CMT), Dehydratase (DH), Enoylreduktase (ER), Acyl Carrier Protein (ACP) und schließlich eine Thioesterase, die für die Abspaltung des fertigen Tubulysins in Form einer freien Säure vom Multienzymkomplex sorgt. Der Insertionsort des Transposons *magellan4* liegt bei der MutT176 bei Basenpaar 54579 innerhalb des Biosynthesegenclusters. Der Insertionsort der Mutante MutT524 liegt nicht auf der uns bekannten Gencluster-Sequenz. Wir postulieren daher, dass der Insertionsort innerhalb eines für eine Acyltransferase codierenden Genes liegt, das downstream vom Tubulysin-Biosynthesegencluster liegt und eine posttranslationale Funktion zur Modifikation des Tubulysins hat.

2. Identifizierung der Anschlusssequenz des Tubulysin-Biosynthese-Genclusters aus *Angiococcus disciformis* An d48

2.1 Identifizierung und Charakterisierung von Cosmiden, die eine überlappende Sequenz downstream zum Tubulysin-Biosynthese-Gencluster tragen

Im vorhergehenden Beispiel wurde beschrieben wie die erste Hälfte des Tubulysin-Biosynthese-Genclusters zusammen mit weiteren an der Biosynthese beteiligten Genen mittels mariner basierender Transposon-Mutagenese und anschließendem Transposon-recovery identifiziert und annotiert wurde. Da innerhalb dieser Sequenz sowohl codierende Gene für Monooxygenasen als auch Acyltransferasen fehlen, sollte noch weitere Sequenz downstream von dieser identifiziert und charakterisiert werden. Innerhalb dieser Sequenz sollten oben angegebene Gene codiert vorliegen, da sie für die Biosynthese des Tubulysins erforderlich sind. Das Biosynthese-Gencluster sollte auf diesem Weg vollständig identifiziert werden.

Hierfür wurde eine Cosmidbank aus *A. disciformis* An d48 mittels Gigapack II XL packaging-kit (Fa. Stratagene) in *E. coli* SURE erstellt. Innerhalb dieser Bank sollten Cosmide identifiziert werden, die einen längeren Überlapp mit dem Tubulysin-Biosynthese-Gencluster downstream von *tubF* besitzen. Dafür wurden aus der bekannten Sequenz des Tubulysin-Genclusters zwei Primerpaare abgeleitet und die PCR-Amplifikate als Sonden für die folgende Hybridisierung der Cosmidbank eingesetzt. Das erste Primerpaar ASTls1A-1B liefert ein 889 bp großes DNA-Fragment und liegt 1 kb vor der *NotI*-Restriktionsschnittstelle in *tubD*. Das zweite Primerpaar ASTls2A-2B generiert ein 700 bp Fragment, das 11 kb upstream im bekannten Cluster-Ende entfernt in *tubC* liegt.

Die PCR wurde bei 54°C Annealing-Temperatur durchgeführt. Durch diese Hybridisierung konnten verschiedene Cosmide identifiziert werden. Mittels PCR- und Restriktionsanalytik wurden sie auf die Größe ihrer Überlappung mit der bekannten Cluster-Sequenz untersucht. Hierfür wurden wiederum die Primerpärchen ASTls1A/B und ASTls2A/B bei einer Annealing-Temperatur von 58°C eingesetzt. Bei der Restriktion wurden verschiedene Enzyme in Einzel- und Doppelrestriktionen eingesetzt.

Da die Cosmide F7 und F13 nach den Restriktionsanalysen einen ähnlich großen Überlapp mit dem ersten Teil des Clusters zeigten, trägt eines dieser Cosmide die nötigen genetischen Informationen, um die unmittelbar clustergebundenen Gene identifizieren zu können.

2.1.1 Southern-Analytik der Cosmide F7 und F13

Für die Identifizierung des richtigen Cosmids wurden zunächst Restriktionsenzyme ausgewählt, die möglichst selten und am Ende der bekannten Genclustersequenz schneiden. Die ausgewählten Enzyme waren *NdeI* und *NsiI*, die an den Positionen 39306 bp bzw. 39430 bp schneiden. Des weiteren schneiden beide Enzyme nur ein weiteres Mal in der bekannten Sequenz. Mit einer generierten Sonde (Primerpaar Tls_{up/down}), die hinter diesen Schnitten direkt am Ende der bekannten Cluster-Sequenz bindet, sollte nun die Cosmid-Genbank „gescreent“ werden. Die Cosmide wurden dafür in verschiedenen Doppelrestriktionsansätzen hydrolisiert und auf einem Agarosegel (0,8 %) aufgetrennt. Für die Doppelrestriktion wurden neben *NdeI* und *NsiI* die Enzyme *BamHI*, *EcoRI*, und *NotI* ausgewählt. Durch die Kombinationen mit *EcoRI* und *NotI* sollte erreicht werden, dass mittels der Hybridisierung ein Fragment identifiziert wird, welches bis an das Ende des jeweiligen

Cosmid-Inserts reicht. Falls dieses Fragment für eine nachfolgende Klonierung zu groß sein sollte, wurde ebenfalls *Bam*HI verwendet, um eventuell kürzere Fragmente zu erhalten. Die Hybridisierung wurde bei 42°C durchgeführt, und es wurde unter höchst stringenten Bedingungen gewaschen (68°C).

Das Ergebnis dieser Analytik war, dass in dem *Eco*RI / *Nde*I-Restriktionsansatz des Cosmids F7 ein 12 kb Fragment detektiert wurde. Dieses Fragment enthält die restliche Sequenz des Tubulysin-Genclusters und reicht bis zum Ende der Insert-Sequenz des Cosmids. Diese Schlussfolgerung wurde aus den Restriktionsanalysen und der charakterisierten Überlappung mit der Tubulysin-Genclustersequenz geschlossen. Das detektierte *Not*I / *Nde*I - Fragment ergab eine Größe von 4,2 kb. Daher muss innerhalb der 12 kb Insertsequenz von F7 - zwischen Anfang (*Nde*I - Restriktionsstelle) und Vektor (*scos*) - mindestens eine weitere *Not*I-Schnittstelle liegen. Somit kann die Anschlusssequenz in kleineren Fragmenten (als *Nde*I / *Not*I - und *Not*I / *Not*I-Fragmente) kloniert und sequenziert werden. Der *Bam*HI / *Nsi*I Doppelrestriktionsansatz lieferte insgesamt fünf Fragmente.

2.1.2. Klonierung der restlichen Tubulysin-Genclustersequenz aus Cosmid F7

Das Cosmid F7 wurde in einem Doppelrestriktionsansatz mit den Restriktionsendonucleasen *Nsi*I und *Eco*RI geschnitten (2 h; 37°C). Nach einer Auftrennung des Restriktionsansatzes mittels 0,8 %igem Agarosegel, wurde die entsprechende Bande aus dem Gel ausgeschnitten und mit dem NucleoSpin kit (Fa. Macherey-Nagel) extrahiert. Das isolierte Fragment wurde mit *Not*I

nachgeschnitten, um zu überprüfen, ob die erzielten Hybridisierungsergebnisse bestätigt werden. Zusätzlich wurde überprüft ob weitere NotI-Erkennungssequenzen innerhalb der 12 kb Anschlusssequenz liegen, um die Anzahl der zu klonierenden Teilfragmente bestimmen zu können. Die Restriktion ergab nach einer Auftrennung auf einem 0,8%igem Agarosegel ein 4,2 (NsiI / NotI-Fragment) und 8 kb großes Fragment (NotI / NotI-Fragment).

Zum einen konnte somit das Hybridisierungsergebnis bestätigt werden und beide Fragmente konnten für eine Klonierung eingesetzt werden. Zum anderen wurde bestätigt, dass es sich um das richtige Fragment handelt, das die Sequenz downstream des Clusters trägt.

Kloniert wurden das 12,2 kb große NsiI / EcoRI - Fragment, sowie die 4,2 kb (NsiI / NotI) und 8 kb (NotI / NotI) großen Fragmente in den Vektor pUC18. Der Vektor wurde für die folgende Ligation mit PstI / EcoRI bzw. PstI / NotI und NotI geschnitten. PstI und NsiI besitzen ein kompatibles Schnittmuster, so dass nach erfolgreicher Klonierung, diese Schnittstellen nicht mehr vorhanden sind. Mittels HindIII bzw. NdeI und EcoRI kann in einer Doppelrestriktion das 12 kb Insert wieder aus dem pUC18 - Derivat herausgeschnitten werden.

Die erhaltenen Klone wurden mittels dieser Restriktionsansätze auf ihre Richtigkeit hin überprüft. Einer dieser Klone (ASpUC12) wurde für eine folgende *in vitro* Transposition mittels GPSTM - 1 Genome Priming System (Fa. New England Biolabs_{Inc}) eingesetzt.

2.1.3. In vitro Transposition mit GPS - 1TM Genome Priming System

Durch das GPS - 1 System sollte mittels einer auf Tn7 basierenden *in vitro* Transposition das klonierte *NsiI* / *NotI* Fragment sequenziert werden. Bei diesem „kit“ wird eine TnsABC Transposase verwendet, die das Transposon (Transprimer™) zufällig in die Zielsequenz inseriert. Durch spezifische Sequenzierprimer (PrimerN / PrimerS), die aus den flankierenden Enden des Transposons „herauslesen“, können die anliegenden Bereiche des DNA-Inserts sequenziert werden. Da das Transposon zufällig in die Zielsequenz inseriert wird, kann durch Sequenzieren einer bestimmten Anzahl generierter Transposon-Mutanten die gesamte Zielsequenz charakterisiert werden.

Durchführung der *in vitro* Transposition

2 µl 10 x GPS-Puffer
+ 1 µl pGPS 1.1 (vermittelt Kan^R)
+ 0,2 µl Ziel - DNA (entsprechen 80 ng ASexp7)
+ 14,8 µl dH₂O

18 µl Ansatz

Dieser Ansatz wird gut durchmischt und mit 1 µl TnsABC Transposase versetzt (wiederum durchmischen). Der gesamten Reaktionsansatz wird für 10 min. bei 37°C inkubiert, damit sich die Transposase vor der eigentlichen Reaktion im Reaktionsansatz vermischt. Nach der Zugabe von 1 µl „start solution“ wird der Reaktionsansatz für eine Stunde bei 37°C inkubiert. In dieser Zeit kommt es zum Strangtransfer des Transposons in die Ziel-DNA. Danach wird die Reaktion durch Inkubation für 10 min. bei 75 °C abgestoppt. Von diesem Ansatz wurden 2 µl in *E.coli* DH10B transformiert und auf Kanamycin-haltigem Medium ausplattiert. Insgesamt sind ca. 2000 Klone nach einer Inkubation über Nacht bei 37°C gewachsen.

Von diesen Klonen wurden 20 daraufhin überprüft, in welchem Verhältnis das Transposon in das Insert oder den Vektor inseriert ist. Dafür wurden diese Klone in einem Doppelrestriktionsansatz mit den Endonukleasen *EcoRI* und *HindIII* hydrolisiert. Die Restriktionsanalyse ergab, dass bei 75% der Klone das Transposon in das Insert inseriert waren. Somit wurden 192 Klone sequenziert, wodurch eine ca. 12 fache Abdeckung der Sequenz erreicht wurde (bei 500 bp Lauflänge je Sequenzierung).

2.2 Sequenzanalyse und Annotieren der 12 kb Anschlusssequenz

Die erhaltene Restsequenz des Tubulysin-Biosynthese-Genclusters ist 12.219 bp lang und hat eine Überlappung mit der vorher identifizierten Sequenz von 133 bp. Sequenzabschnitte, bei denen nur eine einfache Abdeckung vorlag, wurden durch erneute Sequenzierungen spezifischer Klone doppelsträngig sequenziert. In dieser Sequenz ist von Basenpaar 6416 - 6898 eine Acyltransferase (Position 76.787 - 77.545 bp in Gesamtsequenz) codiert. Die anderen identifizierten ORFs haben ebenfalls eine Funktion in der Tubulysin-Biosynthese. Die Gesamtsequenz beträgt somit 82.868 bp.

3. Identifizierung einer Tubulysin überproduzierenden Mutante innerhalb der mariner-Transposon Mutantenbank

Um die Mutanten der Transposonbank auf weitere auffällige Phänotypen gegenüber dem Wildtypen zu untersuchen, wurde eine HPLC-Analytik durchgeführt. Dabei wurde überprüft, ob die Insertion des Transposons in chromosomale Bereiche von

Biosynthese-Genclustern anderer, exprimierter Sekundärstoffe stattgefunden hat. Bei diesen Vergleichen gegenüber einem Extrakt des Wildtypen sollten dann nicht produzierende Mutanten der jeweiligen Metabolite identifiziert werden. Zu diesen Metaboliten gehören das Myxothiazol (Gerth et al. 1980 J Antibiot (Tokyo) 33(12):1474-1479 und Silakowski et al. 1999 J Biol Chem. 274(52):37391-9, Myxochelin (Gerth et al. 1983 J Antibiot (Tokyo) 36(9):1150-6. und Silakowski et al. 2000 Eur J Biochem. 267(21):6476-85) und Angiolam (Kunze et al. 1985, J Antibiot (Tokyo) 38(12):1649-54).

Bei den Auswertungen der 1.200 HPLC-Läufe fielen Extrakte einiger Mutanten auf, bei denen eine erhöhte Myxothiazol-Produktion gemessen werden konnte. Zur Kontrolle der erzielten Ergebnisse wurden 50 ml M7-Medium Kulturen der entsprechenden Mutanten erneut getestet und Zeitkinetiken zur Myxothiazol-Produktion über mehrere Tage gegenüber dem Wildtypen, erstellt. Die Ergebnisse dieser Untersuchungen zeigten eine deutlich erhöhte Produktion des Myxothiazols in den verschiedenen Mutanten bezogen auf den An d48 Wildtypen.

Die Bestimmung des Transposon-Insertionsortes innerhalb der jeweiligen Mutante erfolgte mittels „Transposon-recovery“ und anschließender Sequenzierung der flankierenden Bereiche (siehe 1.4). Die erhaltenen Sequenzen wurden gegen die Genbank auf Homologien zu bekannten Genen hin untersucht und zeigten hohe Ähnlichkeiten zu regulatorischen Elementen / Genen aus bakteriellen Organismen. Aufgrund dieser Ergebnisse wurde die gesamte Mutantenbank auf Tubulysin überproduzierende Mutanten hin untersucht. Dafür wurde der bestehende Toxizitätstest (siehe 1.3) optimiert. Durch mehrfache Verdünnungen des jeweilig eingesetzten Mutanten-Extraktes im Toxizitätstest wird eine

Ausdünnung des Tubulysins erreicht und somit ist die charakteristische Wirkung auf L929 Zellen ab einer bestimmten Verdünnung nicht mehr detektierbar. Anhand dieser Verdünnungsreihen (aus der gesamten Mutantenbank) wurden Mutanten identifiziert, bei denen signifikant höhere Verdünnungen benötigt werden, um keine Wirkung mehr feststellen zu können. Dies bedeutet, dass die jeweiligen Mutante eine erhöhte Produktion des Tubulysins aufweist.

Es konnte die Mutante Mut158 identifiziert werden, die eine vierfach erhöhte Tubulysin D-Produktion zeigte. Dieses Ergebnis wurde sowohl durch die Anzucht der Mutante in 50 ml Kulturen über HPLC-MS Untersuchungen, als auch mehreren Kinetiken mit folgendem, optimiertem Toxizitätstest gegenüber dem Wildtypen gezeigt. Mittels des Toxizitätstest wurde sogar eine achtfache Überproduktion von Tubulysinen ermittelt. Dabei wird die Gesamtwirkung aller Tubulysin-Derivate detektiert und nicht nur die von Tubulysin D. Die Mutante 158 weist somit völlig überraschend noch Überexpression von weiteren Tubulysinen gegenüber dem Wildtypen von *A.disciformis* auf. Dies war in keinster Weise zu erwarten gewesen. Die Ausklonierung des genomischen Bereichs unmittelbar an der Insertionsstelle des Transposons und die Sequenzierung erfolgte wie unter 1.4 beschrieben.

Die Sequenz des betroffenen Gens zeigt hohe Ähnlichkeiten zu einer Proteinkinase (aus *Stigmatella aurantiaca*), wobei der Insertionsort des Transposons die Promotorregion dieses Genes betrifft. Ohne auf diesen Wirkmechanismus festgelegt zu werden, besitzt dieses Gen eine negativ regulatorische Funktion für die Tubulysin-Bildung, weshalb eine Inaktivierung des Gens zu einer erhöhten Produktion führt. Die Gesamtsequenz beinhaltet 2.200

bp, wobei die Proteinkinase von Basenpaar 1.228 - 20 codiert wird und insgesamt 1.209 bp groß ist. Der upstream liegende ORF codiert ein Tubulysin-Biosynthese Protein und ist 933 bp groß.

Primersequenzen:

Sequenzierprimer für das *Himar1* Mini-Transposon *magellan4*

K-388: 5'→3' 5'TGG GAA TCA TTT GAA GGT TGG^{3'} SEQ ID NO: 39

K-389: 3'→5' 5'TGT GTT TTT CTT TGT TAG ACC G^{3'} SEQ ID NO: 40

Primerpaar ASTls1A/B wurde abgeleitet aus *tubD* und ergibt ein 889 bp Fragment

ASTls1A 5'CAC CCG GAC CTG CCT GGA TTC^{3'} SEQ ID NO: 41

ASTls1B 5'TGC TCG GCT GGC GCT ACT CAC^{3'} SEQ ID NO: 42

Primerpaar ASTls2A/B wurde abgeleitet aus *tubC* und ergibt ein 700 bp Fragment

ASTls2A 5'GCT CCC GGG CCA CGT GGT TGA AGA^{3'} SEQ ID NO: 43

ASTls2B 5'CCG CGG GCC GTG GCA GTG GTG TA^{3'} SEQ ID NO: 44

Primerpaar Tls_{up} / Tls_{down} wurde abgeleitet aus *tubF* und ergibt ein 125 bp Fragment

Tls_{up} 5'TGG CAG CCA GCC CGA GC^{3'} SEQ ID NO: 45

Tls_{down} 5'CCG CGG GTG CCC TCT CAT C^{3'} SEQ ID NO: 46

Name des Gens (ob PKS oder NRPS)	Codierende Region in der Sequenz (SEQ ID NO:)	Codierende Region auf dem minus-Strang (SEQ ID NO:)	Größe in bp Größe des abgeleiteten/ übersetzten Proteins in Da	Klassifizierung der Domäne und Position in Nucleotidsequenz oder Protein mit höchster Ähnlichkeit
Funktion SEQ ID NO der Aminosäuresequenz				
ORF16:	3.308 – 1	79.561 – 82.866	3.308 bp	49% Identität, 64% Ähnlichkeit zu Valyl- tRNA Synthetase [<i>Thermotoga maritima</i>] NP 229614
Valyl tRNA-Synthase	(SEQ ID NO: 1)	(SEQ ID NO: 18)	122.434 Da	
SEQ ID NO: 32				
ORF15:	4.706 – 3.453	78.163 – 79.416	1.254 bp	38% Identität, 61% Ähnlichkeit zu "response"
Regulatorische Komponente eines sensorischen Transduktionssystems	(SEQ ID NO: 1)	(SEQ ID NO: 18)	41.337 Da	Regulator CheY Unterfamilie [<i>Synechocystis</i> sp.] NP 440346
SEQ ID NO: 31				
ORF14:	5.719 – 7.164		1.446 bp	29% Identität, 44% Ähnlichkeit zu "response"
Zwei-Komponenten "response" Regulator	(SEQ ID NO: 1)		52.449 Da	Regulator Protein [<i>Mesorhizobium loti</i>] NP 102571
SEQ ID NO: 2				
ORF13:	9.557 – 7.317	73.312 – 75.552	2.241 bp	29% Identität, 46% Ähnlichkeit zu Zwei- Komponenten
Zwei-Komponenten Regulationssystem aus Sensorikase/"response"	(SEQ ID NO: 1)	(SEQ ID NO: 18)	83.814 Da	Regulationssystem aus Sensorikase / "response"
Regulator-Hybrid				Regulator-Hybrid [<i>Agrobacterium</i> <i>tumefaciens</i>] NP 535879
SEQ ID NO: 30				

ORF12: Hitzeschockprotein SEQ ID NO: 29	12.193 - 10.550 (SEQ ID NO: 1)	70.676 - 72.319 (SEQ ID NO: 18)	1.644 bp 58.650 Da	70% Identität, 85% Ähnlichkeit zu Hitzeschock - protein GroEL [<i>Rhodothermus marinus</i>] AAD37976	
ORF11: Tubulysin Biosynthese Protein SEQ ID NO: 3	12.841 - 13.881 (SEQ ID NO: 1)		1.041 bp 37.818 Da	34% Identität, 56% Ähnlichkeit zu hypothetischem Protein [<i>Corynebacterium glutamicum</i>] NP_616546	
ORF10: Tubulysin Biosynthese Protein SEQ ID NO: 28	14.833 - 13.835 (SEQ ID NO: 1)	68.036 - 69.034 (SEQ ID NO: 18)	999 bp 36.441 Da	33% Identität, 47% Ähnlichkeit zu konserviert hypothetischem Protein [<i>Streptomyces coelicolor</i>] NP_631315	
ORF9: Transkriptionsregulator SEQ ID NO: 4	14.942 - 15.586 (SEQ ID NO: 1)		645 bp 23.789 Da	30% Identität, 51% Ähnlichkeit zu Transkriptionsregulator, TetR Familie [<i>Caulobacter crescentus</i>] NP_420005	
ORF8: Integrase SEQ ID NO: 5	15.847 - 16.983 (SEQ ID NO: 1)		1.137 bp 41.614 Da	24% Identität, 40% Ähnlichkeit zu Integrase [<i>Corynebacterium glutamicum</i>] NP_601233	

ORF7: Tubulysin Bioynthese Protein SEQ ID NO: 27	21.154 - 18.809 (SEQ ID NO: 1)	61.715 - 64.016 (SEQ ID NO: 18)	2.346 bp 89.255 Da	33% Identität, 47% Ähnlichkeit zu hypothetischem Protein [<i>Nostoc sp.</i>] NP_490333	
ORF6: Serin/Threonin-Kinase SEQ ID NO: 6	22.366 - 23.532 (SEQ ID NO: 1)		1.167 bp 43.228 Da	32% Identität, 46% Ähnlichkeit zu Serin/Threonin-Kinase Pkn14 [<i>Myxococcus xanthus</i>] AAK64427	
ORF5: Proteinkinase SEQ ID NO: 7	24.591 - 26.513 (SEQ ID NO: 1)		1.923 bp 68.825 Da	42% Identität, 56% Ähnlichkeit zu Proteinkinase [<i>Stigmatella aurantiaca</i>] CAD19078	
ORF4: Adenindeaminase SEQ ID NO: 8	26.597 - 27.517 (SEQ ID NO: 1)		921 bp 33.507 Da	33% Identität, 37% Ähnlichkeit zu Adenindeaminase- verwandtem Protein [<i>Deinococcus radiodurans</i>] NP_285591	
ORF3: Cytosindeaminase SEQ ID NO: 9	29.858 - 30.400 (SEQ ID NO: 1)		543 bp 20.880 Da	24% Identität, 41% Ähnlichkeit zu möglicher Cytosindeaminase [<i>Salmonella typhimurium</i>] NP_462244	
<i>tubA</i> SEQ ID NO: 10	31.220 - 32.392 (SEQ ID NO: 1)		1.173 bp 43.202 Da	Ähnlichkeit zu C-Domänen "core motifs" / Kernmotive C2-C3	

ORF2	33.056 - 32.397 (SEQ ID NO: 1)	49.813 - 50.472 (SEQ ID NO: 18)	660 bp 25.390 Da	29% Identität, 42% Ähnlichkeit zu conserved hypothetical protein [<i>Neurospora crassa</i>] CAD11370	
SEQ ID NO: 26 <i>tubZ</i>	34.195 - 33.074 (SEQ ID NO: 1)	48.674 - 49.795 (SEQ ID NO: 18)	1.122 bp 40.499 Da	39% Identität, 53% Ähnlichkeit zu Lysin- Cyclodeaminase [<i>Streptomyces hygroscopicus</i>] CAA60467	
SEQ ID NO: 25 ORF1	35.422 - 34.205 (SEQ ID NO: 1)	47.447 - 48.664 (SEQ ID NO: 18)	1.218 bp 46.030 Da	40% Identität, 65% Ähnlichkeit zu Anionen transportierende ATPase [<i>Aquifex aeolicus</i>] NP_213468	
SEQ ID NO: 24 <i>tubB</i>	35.522 - 40.147 (SEQ ID NO: 1)		4.626 bp 170.704 Da	C 35.747 - 36.769 A 37.184 - 39.817 NMT 38.369 - 39.730 in A-Domäne zwischen "core motif" / Kernmotiv A8 und A9 PCP 39.818 - 40.069	76 - 416 555 - 1.432 950 - 1.403
SEQ ID NO: 11					1.433 - 1.516

<i>tubC</i> SEQ ID NO: 34	40.144 - 48.021 (SEQ ID NO: 1) 1 - 7.878 (SEQ ID NO: 33)		7.878 bp 289.141 Da	C 40.372-41.397 A 41.824-43.215 PCP 43.216-43.461 C 43.552-44.574 A 44.980-47.631 NMT 46.153-47.547 in A- Domäne zwischen "core motif" / Kernmotiv A8 und A9 PCP 47.632-47.868	77-418 561-1.024 1.025-1.106 1.137-1.477 1.613-2.496 2.004-2.468 2.497-2.575
<i>tubD</i> SEQ ID NO: 12	48.011 - 58.558 (SEQ ID NO: 1)		10.548 bp 383.778 Da	KS 48.011-49.321 AT 49.622-50.584 KR 51.473-52.309 ER 53.066-53.980 ACP 54.158-54.460 HC 54.461-55.870 A 56.000-57.412 PCP 57.413-57.643	1-437 538-858 1.155-1.433 1.686-1.990 2.050-2.150 2.151-2.620 2.664-3.134 3.135-3.211
<i>tubE</i> SEQ ID NO: 35	58.551 - 62.096 (SEQ ID NO: 1) 18.408 - 21.953 (SEQ ID NO: 33)		3.546 bp 130.337 Da	C 58.689-59.714 A 60.156-61.697 PCP 61.698-61.967	47-388 536-1.049 1.050-1.139
<i>tubF</i> SEQ ID NO: 13	62.103 - 70.616 (SEQ ID NO: 1)		8.514 bp 309.369 Da	KS 62.127-63.320 AT 63.711-64.676 KR 64.959-65.882 CMT 65.985-67.061 DH 67.242-67.829 ER 68.247-69.128 PCP 69.360-69.605 TE 69.759-70.586	9-406 537-858 953-1.260 1.295-1.653 1.714-1.909 2.049-2.342 2.420-2.501 2.553-2.828

ORF17: Ähnlichkeit zu Patatin ähnlichem Protein (Lipid Acylhydrolase) SEQ ID NO: 23	71.640 - 70.583 (SEQ ID NO: 1)	11.229 - 12.284 (SEQ ID NO: 18)	1.056 bp 38.371 Da	24% Identität, 40% Ähnlichkeit zu Patatin ähnlichem Protein [<i>Anabaena</i> <i>sp. 90</i>] CAC01602	
ORF18: Ähnlichkeit zu Patatin ähnlichem Protein (Lipid Acylhydrolase) SEQ ID NO: 22	72.786 - 71.731 (SEQ ID NO: 1)	10.083 - 11.138 (SEQ ID NO: 18)	1.056 bp 38.371 Da	24% Identität, 40% Ähnlichkeit zu Patatin ähnlichem Protein [<i>Anabaena</i> <i>sp. 90</i>] CAC01602	
ORF19: Tubulysin Bioynthese Protein SEQ ID NO: 21	75.209 - 74.655 (SEQ ID NO: 1)	7.660 - 8.214 (SEQ ID NO: 18)	555 bp 20.040 Da	31% Identität, 43% Ähnlichkeit zu hypothetischem Protein [<i>Azotobacter vinelandii</i>] ZP_00092207	
ORF20: Tubulysin Bioynthese Protein SEQ ID NO: 14	75.488 - 76.645 (SEQ ID NO: 1)		1.158 bp 43.282 Da	41% Identität, 62% Ähnlichkeit zu hypothetischem Protein [<i>Microbulbifer degradans</i>] ZP_00065421	
<i>tubG</i> : Acyltransferase SEQ ID NO: 15	76.787 - 77.545 (SEQ ID NO: 1)		759 bp 28.039 Da	47% Identität, 53% Ähnlichkeit zu N- hydroxyarylamine O- acetyltransferase [<i>Streptomyces avermitilis</i>] NP_826733	

ORF21: Tubulysin Bioynthese Protein SEQ ID NO: 16	77.769 - 78.695 (SEQ ID NO: 1)		927 bp 33.859 Da	28% Identität, 39% Ähnlichkeit zu konserviert hypothetischem Protein [Xanthomonas axonopodis] NP 641500	
ORF22: Tubulysin Bioynthese Protein SEQ ID NO: 17	79.138 - 80.019 (SEQ ID NO: 1)		882 bp 32.668 Da	37% Identität, 49% Ähnlichkeit zu hypothetischem Protein [Rhizobium eli] NP 659913	
ORF23: Tubulysin Bioynthese Protein SEQ ID NO: 20	81.319 - 80.057 (SEQ ID NO: 1)	1.550 - 2.812 (SEQ ID NO: 18)	1.263 bp 49.133 bp	34% Identität, 52% Ähnlichkeit zu hypothetischem Protein [Nostoc punctiforme] ZP 00109292	
ORF24: Carboxylatreductase SEQ ID NO: 19	82797 - 81.721 (SEQ ID NO: 1)	72 - 1.148 (SEQ ID NO: 18)	1.077 bp 37.621 Da	31% Identität, 44% Ähnlichkeit zu Pyrrolin Carboxylatreductase NosF [Nostoc sp.] AAF17284	
ORF25: Proteinkinase SEQ ID NO: 38	1.228 - 20	973 - 2181 (SEQ ID NO: 36)	1.209 bp 44.079 Da	Tubulysin überproduzierende Mutante	
ORF26: Tubulysin Bioynthese Protein SEQ ID NO: 37	2.157 - 1.225	44 - 976 (SEQ ID NO: 36)	933 bp 33.229 Da	Tubulysin überproduzierende Mutante	

KS: Ketoacylsynthase
AT: Acyltransferase
KR: Ketoreduktase
DH: Dehydratase
ER: Enoylreduktase
ACP: acyl carrier protein / Acyl-Trägerprotein
CMT: C-Methyltransferase
NMT: N-Methyltransferase
A: Adenylierungsdomäne
C: Kondensationsdomäne
PCP: peptidyl carrier protein / Peptidyl-Trägerprotein
TE: Thioesterase

bp: Basenpaare
Da: Dalton

Patentansprüche

1. ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:
 - (i) ss-DNA-Molekül mit einer Sequenz gemäß Figur 1;
 - (ii) ss-DNA-Molekül, das mit einem ss-DNA-Molekül gemäß (i) hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zu jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von dem ss-DNA-Molekül gemäß (i) hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und

- (iii) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines ss-DNA-Moleküls gemäß (i) oder (ii) komplementär ist.

2. ds-DNA-Molekül aus einem ss-DNA-Molekül gemäß Anspruch 1 und einem dazu komplementären Strang.

3. ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:

- (i) ss-DNA-Molekül mit einer Sequenz der Positionen 3.308 bis 1 (ORF 16) der Sequenz gemäß Figur 1;
- (ii) ss-DNA-Molekül mit einer Sequenz der Positionen 4706 bis 3453 (ORF 15) der Sequenz gemäß Figur 1;
- (iii) ss-DNA-Molekül mit einer Sequenz der Positionen 5719 bis 7164 (ORF 14) der Sequenz gemäß Figur 1;
- (iv) ss-DNA-Molekül mit einer Sequenz der Positionen 9557 bis 7317 (ORF 13) der Sequenz gemäß Figur 1;
- (v) ss-DNA-Molekül mit einer Sequenz der Positionen 12193 bis 10550 (ORF 12) der Sequenz gemäß Figur 1;
- (vi) ss-DNA-Molekül mit einer Sequenz der Positionen 12841 bis 13881 (ORF 11) der Sequenz gemäß Figur 1;
- (vii) ss-DNA-Molekül mit einer Sequenz der Positionen 14833 bis 13835 (ORF 10) der Sequenz gemäß Figur 1;
- (viii) ss-DNA-Molekül mit einer Sequenz der Positionen 14942 bis 15586 (ORF 9) der Sequenz gemäß Figur 1;

- (ix) ss-DNA-Molekül mit einer Sequenz der Positionen 15847 bis 16983 (ORF 8) der Sequenz gemäß Figur 1;
- (x) ss-DNA-Molekül mit einer Sequenz der Positionen 21154 bis 18809 (ORF 7) der Sequenz gemäß Figur 1;
- (xi) ss-DNA-Molekül mit einer Sequenz der Positionen 22366 bis 23532 (ORF 6) der Sequenz gemäß Figur 1;
- (xii) ss-DNA-Molekül mit einer Sequenz der Positionen 24591 bis 26513 (ORF 5) der Sequenz gemäß Figur 1;
- (xiii) ss-DNA-Molekül mit einer Sequenz der Positionen 26597 bis 27517 (ORF 4) der Sequenz gemäß Figur 1;
- (xiv) ss-DNA-Molekül mit einer Sequenz der Positionen 29858 bis 30400 (ORF 3) der Sequenz gemäß Figur 1;
- (xv) ss-DNA-Molekül mit einer Sequenz der Positionen 31220 bis 32392 (TubA) der Sequenz gemäß Figur 1;
- (xvi) ss-DNA-Molekül mit einer Sequenz der Positionen 33056 bis 32397 (ORF 2) der Sequenz gemäß Figur 1;
- (xvii) ss-DNA-Molekül mit einer Sequenz der Positionen 34195 bis 33074 (TubZ) der Sequenz gemäß Figur 1;
- (xviii) ss-DNA-Molekül mit einer Sequenz der Positionen 35422 bis 34205 (ORF 1) der Sequenz gemäß Figur 1;
- (xix) ss-DNA-Molekül mit einer Sequenz der Positionen 35522 bis 40147 (TubB) der Sequenz gemäß Figur 1;

- (xx) ss-DNA-Molekül mit einer Sequenz der Positionen 40144 bis 48021 (TubC) der Sequenz gemäß Figur 1;
- (xxi) ss-DNA-Molekül mit einer Sequenz der Positionen 48011 bis 58558 (TubD) der Sequenz gemäß Figur 1;
- (xxii) ss-DNA-Molekül mit einer Sequenz der Positionen 58551 bis 62096 (TubE) der Sequenz gemäß Figur 1;
- (xxiii) ss-DNA-Molekül mit einer Sequenz der Positionen 62103 bis 70616 (TubF) der Sequenz gemäß Figur 1;
- (xxiv) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii) oder (xxiii) unter stringenten Bedingungen hybridisierbar ist und insbesondere dieselbe Anzahl von Basen aufweist; und
- (xxv) ss-DNA-Molekül, das mit einem ss-DNA-Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii) oder (xxiii) hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zu jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von diesem ss-DNA-Molekül hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und

(xxvi) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines Moleküls gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv) oder (xxv) komplementär ist.

4. ds-DNA-Molekül aus einem ss-DNA-Molekül gemäß Anspruch 3 und einem dazu komplementären Strang.

5. ss-DNA-Molekül, ausgewählt aus der folgenden Gruppe:

- (i) ss-DNA-Molekül mit einer Sequenz der Positionen 35747 bis 36769 (Domäne C des tubB-Gens) der Sequenz gemäß Figur 1;
- (ii) ss-DNA-Molekül mit einer Sequenz der Positionen 37184 bis 39817 (Domäne A des tubB-Gens) der Sequenz gemäß Figur 1;
- (iii) ss-DNA-Molekül mit einer Sequenz der Positionen 38369 bis 39730 (Domäne NMT des tubB-Gens) der Sequenz gemäß Figur 1;
- (iv) ss-DNA-Molekül mit einer Sequenz der Positionen 39818 bis 40069 (Domäne PCP des tubB-Gens) der Sequenz gemäß Figur 1;
- (v) ss-DNA-Molekül mit einer Sequenz der Positionen 40372 bis 41397 (Domäne C des tubC-Gens) der Sequenz gemäß Figur 1;

- (vi) ss-DNA-Molekül mit einer Sequenz der Positionen 41824 bis 43215 (Domäne A des tubC-Gens) der Sequenz gemäß Figur 1;
- (vii) ss-DNA-Molekül mit einer Sequenz der Positionen 43216 bis 43461 (Domäne PCP des tubC-Gens) der Sequenz gemäß Figur 1;
- (viii) ss-DNA-Molekül mit einer Sequenz der Positionen 43552 bis 44574 (Domäne C des tubC-Gens) der Sequenz gemäß Figur 1;
- (ix) ss-DNA-Molekül mit einer Sequenz der Positionen 44980 bis 47631 (Domäne A des tubC-Gens) der Sequenz gemäß Figur 1;
- (x) ss-DNA-Molekül mit einer Sequenz der Positionen 46153 bis 47547 (Domäne NMT des tubC-Gens) der Sequenz gemäß Figur 1;
- (xi) ss-DNA-Molekül mit einer Sequenz der Positionen 47632 bis 47868 (Domäne PCP des tubC-Gens) der Sequenz gemäß Figur 1;
- (xii) ss-DNA-Molekül mit einer Sequenz der Positionen 48011 bis 49321 (Domäne KS des tubD-Gens) der Sequenz gemäß Figur 1;
- (xiii) ss-DNA-Molekül mit einer Sequenz der Positionen 49622 bis 50584 (Domäne AT des tubD-Gens) der Sequenz gemäß Figur 1;

- (xiv) ss-DNA-Molekül mit einer Sequenz der Positionen 51473 bis 52309 (Domäne KR des tubD-Gens) der Sequenz gemäß Figur 1;
- (xv) ss-DNA-Molekül mit einer Sequenz der Positionen 53066 bis 53980 (Domäne ER des tubD-Gens) der Sequenz gemäß Figur 1;
- (xvi) ss-DNA-Molekül mit einer Sequenz der Positionen 54158 bis 54460 (Domäne ACP des tubD-Gens) der Sequenz gemäß Figur 1;
- (xvii) ss-DNA-Molekül mit einer Sequenz der Positionen 54461 bis 55870 (Domäne HC des tubD-Gens) der Sequenz gemäß Figur 1;
- (xviii) ss-DNA-Molekül mit einer Sequenz der Positionen 56000 bis 57412 (Domäne A des tubD-Gens) der Sequenz gemäß Figur 1;
- (xix) ss-DNA-Molekül mit einer Sequenz der Positionen 57413 bis 57643 (Domäne PCP des tubD-Gens) der Sequenz gemäß Figur 1;
- (xx) ss-DNA-Molekül mit einer Sequenz der Positionen 58689 bis 59714 (Domäne C des tubE-Gens) der Sequenz gemäß Figur 1;
- (xxi) ss-DNA-Molekül mit einer Sequenz der Positionen 60156 bis 61697 (Domäne A des tubE-Gens) der Sequenz gemäß Figur 1;

- (xxii) ss-DNA-Molekül mit einer Sequenz der Positionen 61698 bis 61967 (Domäne PCP des tubE-Gens) der Sequenz gemäß Figur 1;
- (xxiii) ss-DNA-Molekül mit einer Sequenz der Positionen 62127 bis 63320 (Domäne KS des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxiv) ss-DNA-Molekül mit einer Sequenz der Positionen 63711 bis 64676 (Domäne AT des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxv) ss-DNA-Molekül mit einer Sequenz der Positionen 64959 bis 65882 (Domäne KR des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxvi) ss-DNA-Molekül mit einer Sequenz der Positionen 65985 bis 67061 (Domäne CMT des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxvii) ss-DNA-Molekül mit einer Sequenz der Positionen 67242 bis 67829 (Domäne DH des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxviii) ss-DNA-Molekül mit einer Sequenz der Positionen 68247 bis 69128 (Domäne ER des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxix) ss-DNA-Molekül mit einer Sequenz der Positionen 69360 bis 69605 (Domäne PCP des tubF-Gens) der Sequenz gemäß Figur 1;

- (xxx) ss-DNA-Molekül mit einer Sequenz der Positionen 69759 bis 70586 (Domäne TE des tubF-Gens) der Sequenz gemäß Figur 1;
- (xxxi) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix) oder (xxx) unter stringenten Bedingungen hybridisierbar ist und insbesondere dieselbe Anzahl von Basen aufweist;
- (xxxii) ss-DNA-Molekül, das mit einem Molekül gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix) oder (xxx) hinsichtlich seiner Nucleotid-Anzahl oder seiner Nucleotid-Sequenz zur jeweils 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 oder 100% homolog ist, jedoch von diesem ss-DNA-Molekül hinsichtlich seiner Nucleotid-Anzahl und/oder seiner Nucleotid-Sequenz in mindestens einem Nucleotid abweicht; und
- (xxxiii) ss-DNA-Molekül mit einer Sequenz, die zur Sequenz eines Moleküls gemäß (i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x), (xi), (xii), (xiii), (xiv), (xv), (xvi), (xvii), (xviii), (xix), (xx), (xxi), (xxii), (xxiii), (xxiv), (xxv), (xxvi), (xxvii), (xxviii), (xxix), (xxx), (xxxi) oder (xxxii) komplementär ist.

6. ds-DNA Molekül aus einem ss-DNA-Molekül gemäß Anspruch 5 und einem dazu komplementärem Strang.
7. Varianten oder Mutanten, die aus einer Substitution, Insertion oder Deletion von Nucleotiden oder einer Inversion von Nucleotid-Segmenten eines ss-DNA-Moleküls gemäß Anspruch 1, 3 oder 5 oder eines ds-DNA-Moleküls gemäß Anspruch 2, 4 oder 6 resultieren, wobei diese Varianten und Mutanten Enzym-Varianten oder Enzym-Mutanten für die Produktion von Sekundärstoff(en) mit für Tubulysine charakteristischen Eigenschaften codieren.
8. RNA
 - (a) mit einer Sequenz entsprechend der eines ss-DNA-Moleküls gemäß einem der Ansprüche 1, 3, 5 oder 7 oder
 - (b) mit einer Sequenz einer RNA gemäß (a), aber in Gegenrichtung (anti-sense), oder
 - (c) mit einer Sequenz einer RNA gemäß (a) oder (b) und mit einem dazu komplementären Strang,jeweils gegebenenfalls als Element eines rekombinanten Vektors.
9. Rekombinanter Vektor, insbesondere Expressionsvektor mit einem DNA-Molekül gemäß einem der Ansprüche 1 bis 7.
10. Zelle, insbesondere zur Expression, in die ein DNA-Molekül gemäß einem der vorhergehenden Ansprüche oder einen Vektor gemäß Anspruch 8 oder 9 integriert ist.
11. Zelle nach Anspruch 10, wobei sich die Zelle von kultivierbaren Bakterien, insbesondere Myxobakterien, vorzugsweise Angiococcus, insbesondere A. disciformis, Archangium, insbesondere A. gephyra, Escherichia coli, Pseudomonaden oder Actinomyceten herleitet.

12. Verwendung eines Vektors gemäß Anspruch 8 oder 9 für die Transformation von Zellen oder Organismen zur transienten oder permanenten Expression eines oder mehrerer Proteine (Expressionsprodukt(e)), das (die) durch eine DNA (ssDNA oder dsDNA) des Vektors codiert wird (werden).
13. Verwendung einer Zelle gemäß Anspruch 10 oder 11 zur enzymatischen Biosynthese, Mutasynthese oder Partial-Synthese eines Tubulysins, insbesondere Tubulysin A, B, C, D, E und/oder F.
14. Expressionsprodukt eines DNA-Moleküls gemäß einem der Ansprüche 1 bis 7 oder eines Vektors gemäß Anspruch 8 oder 9 oder einer Zelle gemäß Anspruch 10 oder 11.
15. Polynucleotid enthaltend eine Sequenz wie in SEQ ID NO: 1, 18, 33 oder 36 definiert, oder ein Fragment davon.
16. Polynucleotid nach Anspruch 15, wobei das Fragment eine im Sequenzprotokoll definierte CDS ist.
17. Vektor enthaltend ein Polynucleotid nach Anspruch 15 oder 16.
18. Zelle enthaltend einen Vektor nach Anspruch 17.
19. Polypeptid enthaltend wenigstens eine Sequenz wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder ein Fragment und/oder Derivat davon.

20. Verwendung wenigstens eines Polynucleotids wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens eines Fragments davon und/oder wenigstens eines Polypeptids wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens eines Fragments davon zur Herstellung einer pharmazeutischen Zusammensetzung zur Behandlung von unerwünschtem Zellwachstums oder unerwünschter Zellvermehrung in einem Individuum.

21. Verwendung nach Anspruch 20, wobei das unerwünschte Zellwachstum oder die unerwünschte Zellvermehrung ein Tumor ist.

22. Verwendung nach Anspruch 20, wobei das unerwünschte Zellwachstum eine pathogene Infektion ist.

23. Verwendung nach Anspruch 22, wobei die pathogene Infektion eine Mykose, Malaria oder eine parasitäre Erkrankung ist.

24. Pharmazeutische Zusammensetzung, enthaltend wenigstens ein Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens ein Fragment davon und/oder wenigstens ein Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens ein Fragment davon.

25. Pharmazeutische Zusammensetzung nach Anspruch 24, die weiterhin wenigstens einen pharmazeutisch verträglichen Trägerstoff enthält.

26. Verfahren zur Herstellung von Tubulysinen und Tubulysin-Biosynthese Proteinen, umfassend die Schritte:

- (a) exprimieren von wenigstens einem Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder

wenigstens einem Fragment davon und/oder wenigstens einem Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens einem Fragment davon, und

(b) aufreinigen der Expressionsprodukte.

27. Verfahren nach Anspruch 26, wobei die Expression in prokaryotischen oder eukaryotischen Zellen und/oder durch in vitro Expression erfolgt.

28. Verfahren zum Auffinden von Genen, die an der Biosynthese von Tubulysinen beteiligt sind, umfassend die Schritte:

(a) hybridisieren von wenigstens einem Polynucleotid wie in SEQ ID NO: 1, 18, 33 und/oder 36 definiert und/oder wenigstens einem Fragment davon mit DNA, RNA, und/oder cDNA einer Spezies, die nicht mit *Angiococcus disciformis* identisch ist, und

(b) isolieren und charakterisieren der hybridisierten DNA, RNA, und/oder cDNA.

29. Kit zur Herstellung von Tubulysinen, enthaltend:

(a) wenigstens ein Polynucleotid nach Anspruch 15 oder 16 und/oder wenigstens einen Vektor nach Anspruch 17 oder

(b) geeignete Medien und Puffer zur Vermehrung von Zellen, die eine Expression des Polynucleotids und/oder Vektors erlauben und

(c) geeignete Mittel zum Aufreinigen des/der Expressionsprodukte(s).

30. Verwendung einer Zusammensetzung enthaltend wenigstens ein Polypeptid wie in SEQ ID NO: 2 bis 17, 19 bis 32, 34, 35, 37 und/oder 38 definiert und/oder wenigstens ein biologisch aktives Fragment oder Derivat davon als Desinfektionsmittel.

31. Verwendung nach Anspruch 30, wobei die Zusammensetzung flüssig oder pulverförmig ist.

32. Desinfektionsmittel wie in Anspruch 30 oder 31 definiert.

Figur 1

SEQUENCE LISTING

<110> Gesellschaft fuer Biotechnologische Forschung mbH

<120> Tubulysin-Biosynthesegene

<130> 13289

<160> 1

<170> PatentIn version 3.1

<210> 1

<211> 70782

<212> DNA

<213> *Angiococcus disciformis*

<400> 1

gcgggcgccct tcttcaccgc gggggcctcg gcctcctcgg tggtcggggc gggcgccagc	60
ttgtcgccgt gcaccttgcc gggagcggtc gccttcttct tcgaggcgga gcgcacgggc	120
tcgccctgcg tcacctcgcg cgtgggcagc atgtcccgt gggccttcac cgggccggac	180
tcacctcct ccgcctcggc gccctcgcc ggcgctcgc ggggtgcgc accctccttg	240
gcccgggtga actcgcgcat cgcgtcgtcc accagcccca tggtcagtga cgccacggcg	300
aggatcatggt ggtcggtatg cgacaggccc tcttcgtgc cctcgcgag cttcgcgagc	360
gcgtcctgca cctgcgtgtc cccggcctcg gacgcgggtc cggctcgcg cagttcacc	420
ttcagctcct ggcccaggtc cagccgcgcc ttctcttct gattcggggc gaccttcacc	480
tggggcgggg cggaggggga ggcacctct tccgcgggtg ggggtgctgt ctcgggagtg	540
gtgctctcgg acggcgggct ctctggcatg gcgggctccg gggcaatccg ctgcagggtg	600
tcctgaagct tggccttgcg ctcttcagc tcctcgacgc gggcgcggtc cttctccacc	660
acgtccgggg gcgccttggc caggaagttg gggttctcca gcttgcgag cagccgggc	720
atctcctgct cggcgcgggc aatctccttc ttcaaccgct cgcgctcgc gtccaggtag	780
acgaggcccg ccagcgggac gtagatctcc aggttgagc ccacgaaggc ggcggcctgc	840
ggcggttg cggccgggc tcccaccgtc acctcgaca ggccggccag cggcatgagg	900
tagccgcgcc agcgtccag cagctcgcgc gtgcgcacgt ccgcgtctg caccaccgc	960
ttcaccttg tggcgggga cagggtgtc tcgcccgga tgggtgcgag gccctcgatg	1020
gcggcgatga cggcgccat ctgccttcc gccgctcgt ccacgtgcgc cgcgtccggc	1080
tccgggtacg gggcaatcat gatgctgtc gtgcggcgg ccatcggcag cttctgccag	1140
atctcctcgg tgatgaacgg catgaacgg tgcagcagc gcaggatgc gtccaggcag	1200
tacacgagca cccggcgcg ggagtcctc gccgccgct cctcgccgta cagcgagccc	1260
ttcgccagct cgatgtacca gtgcgagaac tccgccaga ggaactggta gagcgtggag	1320
gcggcctcgc cgaagctgta cgactccagc agcgcgcgc cctccgtggt ggcgcgctgg	1380

2/38

aggcgggaga gaatccagcg gtccgccagc gtcagtgtgc tccgggtccag cgggcccggcc 1440
tccatccgga agtcgcccatt gttcatcagg gcgaagcggc tggcggtcca cagcttggtg 1500
ccgaaggcct tgtagccctc gagccggtcc atcgacagct tgatgtcgcg gccctgctgg 1560
gtgagcgagg cgagcgtgaa gcgcagcgcg tccgcgccga acggcgccat gccctgcggg 1620
aaccggttct tcagcgtcgg ctggagcttg tcggccgggg cgccgaggac gatgtccagg 1680
gggtcgatga cgttccccct cgtcttggaac atcttctggc ccttctcgtc gcgcaccatc 1740
gcgtgcagggt acacggtgcg gaagggcaca tcccccatga agtggtatgcc catcatcatc 1800
atccggggcga cccagaagaa gatgatgtcg tggcccgtct ccatgacgga cgtcgggtag 1860
aaggtcttca gctccggcgt ctcacgcggc cagcccagcg tggagaacgg ccacaggccg 1920
gacgagaacc aggtgtccag cacgtccggg tcctggatga aggacgcgcc gccgcacttc 1980
gggcaggact tcggtgctc gcgcgcgacg atgggtccg cgcgcgcgaa gtccacccccg 2040
cccaccttca ccgtgggggc gtccagcggc aggtccgtgt cgtcgccctg ccgcgggctg 2100
cacgaggtgc agtagtacgc gggaaatctgg tggccccacc acagctggcg gctgacgcac 2160
cagtcgtgga tgttgcgcat ccagtggaa aacgtgttcg tccaggactc ggggacgaac 2220
ttcgtgcggc cctgctccac cgcctcgatg gccggcttcg ccagcgggtc aatcttgatg 2280
aaccactgcg gagacaggcg tggctccacc accgtggcg agcgtggca ggtgcccacg 2340
ttcagcatgt ggggctcttc cttctccagc agcccctgct ccgtgaggtc cgccagcacc 2400
tgcttgcgcg cctcgaagcg atccatgccg gcgtacttgc cggctctcctt cgtcatccgg 2460
gccgcttcgt ccaggatggt gagcatgggc agcttggtcc gcaggcccgt ctggtagtcg 2520
ttgaagtcgt gcgcggcgt caccttcacc acgccggtgc cgaacttcgg gtccaccagc 2580
tccgcgtccg cgatgatggg aatctcgcg tgggtcagcg gcagcttcac gctcttgccg 2640
gccagcccca ggtagcgtc gtccctccggg tggatggcca ccgcggtgtc gcccgatc 2700
gtctccgggc gtgtggtggc cacggtgagc gtgcggtcgc tgtccttgac ggggtagcgg 2760
atgtgccaga tggagccctt cttcgactcc tcgtgtcca cctccaagtc gctgagggcg 2820
gtgcggcacg aggggcacca gttgatgagc ttctgggccc ggtacatcag gccctcttcg 2880
tacaggcgca cgaagacctc gcgcacggcg gcggaggact gctcgtccat ggtgaagcgc 2940
tcgcggctcc agtcacgca ggcgcccagg aagcgggtgt gctcgccgat gcgggcgcgg 3000
taattgccct tccactccca gacgcgtcc aggaaggcgg cgcgcccgag gtcgtgccgg 3060
ctcttgccct cggctctctt cagctccttc tccaccacca tctgcgtggc gatgccggcg 3120
tgggtccgtg cggggagcca cagggtgttg aagccgtca tccgcttcca gcgcgcgagg 3180
atgtcctgga tgggtggcgg gagcgcgtgg ccgatgtgga ggctgcccg caggttgggc 3240

3/38

ggcggcagca cgatggagaa ggcgggcttg tcggaggctg cctcggcgcg gaagtagttc 3300
cgctccatcc agcaggcgta ccagcgggcc tcgacctcgg tgggctcgta ggccttgagc 3360
agttcagtgg tgtcgggtcat tgcggacggc cagccccggg ggaaccgggg ccggcgggaa 3420
ggtcttgag agaggttaag gcgaggaacg cgtcagtgtc gcgtctcccg gtccttgatg 3480
agccgctcga gctcctcacg gatgatggtc tccgccagct gcggtaccac ctcccaggca 3540
atcttctcga tgacctcgg ggaggccttc gagagcgctt cgcgcagcag cgcctcgcca 3600
ccatccgcg cggtcggacg ggcggtggcc ggagcggcg tcggcgccac cgcggtcgga 3660
gcggccgtcg gggggccgcc catgtccagg gagatctcct cggcaccgct gggctcgggc 3720
agcgagtcct cgatgctgat ggaaggctgg gctgcctcgg gcgcgcgct gggagcgccc 3780
aggccgaacg ggtccctcgc gcgggcccgc ggctgggtcg cgcaccccg cgcgggaggg 3840
agcgcgcgcg cgcgcgccgg aggacgcggg aagcccggtt ggctgccttg gggaatcccc 3900
ggacggggcca tgccggcgcc ggcaccccg gaggcgcggg cagccccgga cgcggcgga 3960
tgcccgcgcg ggacgcaccc ggcggaggcg gcacgcgccg gacgcgcggc ggcgcgccga 4020
ggcgcatgc ccgggcccgg aggacgcggc gcgcgccggg caccggcgcg cggcattcca 4080
ggccccggcg ggcgggccat gcccggtccc ggcggcattc caggccccgg aggacgcggc 4140
gcgcgccggg caccggagg cggcattcca ggccccggg gacgggcat gcccgcccc 4200
ggcggcattc caggccccgg aggacgcggc gcgcgccggg caccggcgcg catgccccgg 4260
ccccgaggac gcggcacgcc cggctgggtg cctgcggaa tgcccgagc ggcggccgtg 4320
ggcgcgcccc gctgcgggcc cggaggcacc ggcgcgggag tgcccggtg ggcgggtggc 4380
ggcgcgacct gcgtggccgc ggaggccggc atggtgttg acttctgccc caccagcgcc 4440
ttcaccttgt cgagcagcac ctggtctcgc aagggttgg cgatgtggtc atcggcgcgg 4500
gcggcgcggg cgcggttctc gtcgaaggcc tcgaaggctc ccgccagcag caccacgggg 4560
atgccctggg tgcccggtc gctcttcagg gcctcgcaga cctcgtagcc gctcttgccc 4620
ggcatcatca catcggcgag gacgacgtcc gggcgagct cgcgggtacg ggagatggcg 4680
tccagcccgt tgctcaccgc ggtcacctga aagtcctcgg tcggaagat catccaatc 4740
accttgcgga tggtagcga gtcacggcg accagcagat tcttgggcat cggttcgggc 4800
ctcggggggc gtccacccc cctgaattcg ttggagattc ctgcccgaa aaaccaggt 4860
cgcgccctg tatcaagcg ggcagcttac gtttcgcgt ccagcacgat caagaaaaca 4920
tgcgctgcag gtccaggaac agcacggagc ctgtcaacct gggggcacgg aaggctcaa 4980
tggtggtctc cggctcgaag tgctgcagca cgccgagcac cgggtgggc gtcagcccca 5040
cgttccggcc ggccagctcc gtgaggatga agagggcctc gggtagcggg ggggagccca 5100
gcagcgcggg gaccgagcag atgggccaca gcacctgggc gtgcgggaag atgcccga 5160

4/38

ccggccccgt ctgcacgggc aggacgctga aggcctcccc tcgcgatata acctgggaga 5220
cgaagcccag cgggacgccg aacaggcgcc cctgcgattc gaagacgagc gccttgctcg 5280
gcgtggactc ggcccagtgg acggggcgca gggccgcggg cgcggcacga gggcccaccc 5340
ggtggggcag cgctcggcg atgagctcca ggtagagccg ctcttctgtc agcaccgcgc 5400
cccggctcag gggggccagc gagtccgcca gcccgggggg caggaggaag aacggcgccc 5460
gcgccacgtc ggcgacctcc accacggagc gcacccgcac ggccagggtg ggactcacgt 5520
cgagcaccac caccatgccc ggcccctcct cgggcaagcc gccgagcagg gccgacaggt 5580
ccttcacctc cagcacgcca cgcaggctgg agccgttggc gccgggcatg gccacttcca 5640
tgacggaggt ggcctccacc gcgtagcgcg tctctcctgc ctccacgaga aggcagagcc 5700
ggcgtccgct ttcgaagggt gacacggccc taccctagca gccttcctcg ccttggtggt 5760
aagccatcga gccagatggc gcgcatactc ctcgtcgacg acgaaaagat cgcccgaccc 5820
ctctacggcg actacctcac ggccgtgggg caccgcgtca cggcgggtgg ctcgctccag 5880
gacgcgcggg aggcgctggc cggggaccgc ttcgacgcgg tggtagcga cctcatcctc 5940
cccggcgggc acggcatgga ggtgctgcgc cactgctgcg agcgccaccc cggcgtggag 6000
gtggtggtca tcaccggcct ggagaagggt gaccccgccg tgcgcgccat caagagcggc 6060
gcggccgagt acctcgtcaa gccggtggcc cccgaggccc tccagcatgc cgtccgccgt 6120
gccctacca cgcgagacct cctccaggag aacgcctcgc tgcgcgggca cgtcgccatg 6180
ctggaggccg gccagcgcat cgccaccacg ctggagcgcg agaagctggc ctccgccacc 6240
acgagcgcgc tggagtccat ggctgcgc agcgccgtgg tgctgctgga gcgcgacggc 6300
gacaagggcc tgcggctcca gggcgctcgc ggactgcctc cgcggcacga gccggagctg 6360
gtggccaccc tcgtcgagcg cctctcgaac gcgcgcgcgc cccgggagct ggaagggtg 6420
gacgccgcct tcccgcgcgt cctctccatc cccgccgtgg agggcaccac ggtgctgggc 6480
cacgcggtgc tcttcttcga cggcgccgag ccggagtgga ccggcgagac ggccggctac 6540
ctggtgcgca actgggcact ggcgctgcgc aacctcgggc gcttcgcggc ggtggaggac 6600
ctggcgtagc tcgacgacct caccgcctc ttcaacacgc gctacctgca cctggtgctg 6660
gaccgcgagg tccaggactc cgtccagacg cagcgccctc tcagcctgct gttcctggac 6720
ctggaccact tcaagtccat caacgacacg caccggccacc tcgtgggctc gcggctgttg 6780
gtggagacgg cccgggtggt gaagggtgc gtgagagacc acgacgtggt ggcccgtac 6840
ggcggcgacg agtacgtggt gctcctgcgc aacaccgact ccggcggcgc gctcaagggtg 6900
gcggagcgca tccgcgcac catggagacg cacctgttcc tcgcgcgca ggggtgtcg 6960
ctgaagctca gcacgtgcat cggcgtggcc agcttccccg agcacgccg ggacaaggcc 7020

5/38

acgctgctgg acctggcgga ccgggccatg taccggggca agcgtggctc ccggaatgtc 7080
gtctacatgg cgccctgga tctggagccg ctgccggcgg agcggcgcca gggcggcagc 7140
ggccccggcg gctccagcac gtagcggctc agcggcgagc gttgcccacg gtgaggcgct 7200
tggtcggggg ggtctccacc tcggcgcggtg ctcccccgcc ctcttgccc accgctcca 7260
cctcctcgtg cggggcaccg gtggtgatga gccggtactt ctccgaggcc tcgcggtcat 7320
gctgggcctg ctcggggtcc aaccgggcaa tcagctgcca gaagagggcg gcgtgctcgc 7380
gctcctcgtc catgacgtga cggaggatgg ccttggcctc ctcgttgtcg gtggcatcga 7440
tgtgggcggc gtagagggtg atggcgcca gctccgcctc gatgttcagc cggatggagc 7500
gggccagctc ggagtcggtc atcttcggg gccaccagca gtggaacggg ttggtctgcg 7560
gcatgaaggc cctcccgagg ccggtggtgg cgtccacgtt ctggacgggg cgccagggga 7620
cgcgaggagc atccgcgcg ccagggggcc gggtaacgg aatccctccg cccgcgttgc 7680
gtagataccg ctgcctccgc catgctcggg ccatggaaac accgcgccc ctctcccagt 7740
tgctccaggc cctggaggcg ggagatctgc cggccgcgcg cgcggcgcg gcggctctgc 7800
aacgtgcggg cgcgcactcg gcacagctgg cggcggaggt gctgcacgag ctgcggcagc 7860
ccctgctggg cgtgaaggcc tacgcgcagc tgctcgcgga ggacgggtgg cccaccggcc 7920
cgctgcggct gctgctggcg cagggtggagc ggatggaaca gatcgtctcc gactacatcc 7980
gcctcgccag cgagcggcac gcgcgcagc agcgcctgtc gctggcgggc cccatctggg 8040
cggcggcgaa gctgttcagc gtcaaccgg actcggcgcg catgtctctg gaggtggagg 8100
cgcccgagga catcaccatc cagggcaacg cgcggctcat cgagcagctc acgctgaacc 8160
tgctgaacaa cgcgcgcgac gccatggccg gccgcggcg cgtgaagggtg gtgctctcgc 8220
gcgagggcgc ctctccgggtg ctgtacgtgg cggactgggg cccgggcac cccgtggagc 8280
tgcgcgagcg catcttcgag ccctacgtca ccgccaacaa gcgcggcacg ggactggggc 8340
tgggcggtgtg caagcgcac cccaggagc acaacgcgac gattggcctc gcggcgccgg 8400
gcgccattcg cgacgtgccc ccgcccggca ccgtgttccg ggtactcttc cccgccacgg 8460
acgcgcggcc gcagctgcgc aagcggctgc tgggtgggtga cgacgagacc atcatccga 8520
tggtcttccg cgacctgatg ggcaaggagt gcgaggtcat cgaggcggcc agcggcgagg 8580
aagcgtgga tttgctgcgc caggcgccag tggacctcat cgtcacggac aagaacctgc 8640
cgggcctgtc cgggctggag ctggcgcagc aggcggcgcg gctctactcc aactcgcgcg 8700
tcatcctgat gacgggctac ccgtcgtgg tgacgacgca gcaggcgctg gagctggggc 8760
tggtggacta cctgctcaag cccttcgacg acatccggca ggtgcgcggg ctgttgcgca 8820
cgacgctgtc ctcgagagcg ccggtgcctc cgggtgggtgg ccccgggggc gaagtgcggc 8880
gggtggacgt gctggaggac aaccgacga cggcgcggt catctccgag gccctggaca 8940

6/38

tgctggggct	ggaggcgcg	gtgctgccct	cgacggagct	gatggcgatg	gcgacgccgg	9000
tgggcgtggt	ggtgagctgg	gacttcgcgc	cggcctatgg	gcgcaaggcg	ctggagctgg	9060
ccaaggcgct	ggcccagggc	gcccccttcg	tcgtgctggc	cgagcacctc	accatggaga	9120
cggcgctgga	gtcgctgcgc	gccgggtgcc	ccgcgtgcct	tccgaagctg	ctgtccgata	9180
ccacggcgct	cagccgtgag	ctgagccgcg	ccttcaagag	agaggtccca	tgagactcgc	9240
gctcctgtcc	gtcctgctct	tcgcggtggc	ctgttcctcg	gacaagcccc	cgccgacgga	9300
gcggccggac	gggggcgcg	cggaggactc	cgggacgggc	ggagaggacg	cgggccccgac	9360
ggaggacgga	ggcacggagc	cggatggtgg	ctccggctcc	gatgcgggca	cggatgcggg	9420
agcgggcgac	gcgggacgg	tggatggcgg	aagctgtgag	gcgcccgcgg	ccgagccccg	9480
tacggagctg	gccacgcgac	tgaacacgcc	ccgcccggctc	gcggtggaca	cgacggacat	9540
ctacatctcc	gagtcacact	ccctgaaccc	ccagcagccg	agtcggggac	cgggacaggt	9600
tctgaggctg	ccccgcgcgg	gagggtccacc	cacggttctg	gccacgggct	tccgcgcacc	9660
agacgccatc	gccgtggatg	ggacgagcgt	ctacgtgctc	gacctggacg	ggctctggcg	9720
ggtggacaag	gcaaccggga	agcggggcga	cctgcccata	gacgcgacgg	atcgcaacgt	9780
caccgtcggt	ggtaccgagg	tgctccgcgc	cacgctcgcc	ggacgcgacg	tgctggtggt	9840
cgctacgggg	cggcagtcgc	tgggtccgggt	ggacaccgat	ggcggcaact	cccaggttct	9900
ctacaccggt	cccggcggtc	ccctggtgcg	tggcgcgcg	gtggtggact	cggacgtgtg	9960
gttcctcgtc	agcgcgggga	cgggagccgg	aaacagctcg	ggtctgtacc	gcgtcccgtc	10020
ggacggcagc	gcgcccgcgg	agcggcggg	cgccaccata	ctccaaggca	attcactgga	10080
ggtgaccccc	acgcacttcc	tcgtcaccga	ggcggtggg	ggtaccggcc	gggtgttgcg	10140
gctgccccgc	gctggaggca	ccgccgaggt	gctggccgac	ggtctccagg	gcccgtgggt	10200
cccgttgga	ctgaatggga	ccatctactt	caaggagtca	cgtgccgacg	gcgcggactt	10260
cctgcgccgc	gtgcgcacgt	gcgcgctggg	gacgtcggat	ccggtggggc	ccccgggcac	10320
gggaccaggc	gggctcatcg	tggacggcag	caccctcctc	tacacgtcgc	aggagagcgg	10380
caccggcggc	gccgtgggcc	gcgtgccctg	acgccctcct	ccacccccga	aaagaccacg	10440
gccccggcgc	ctcgcgagga	ggcatccggg	gccgcgtcct	gcatacgtgg	aacacctcac	10500
cggctcggcg	ggcccgaggc	ctcatcggcc	ccggaccgcc	gtgagcggct	cagtagtcca	10560
tgctcgtcgc	gccgtagtcc	ggcatcccgg	cgccgccgcc	gttcttgccc	ttcttcttgg	10620
cggggcggtc	ggcgatcatc	gcctcgggtg	tcagcagcag	cgaggccacg	gacgcggcgt	10680
tctgcagcgc	ggtgcgctcg	accttcgctg	ggtcgatgac	gccggccttc	tccaggtcct	10740
cgtagacctc	ggtgcggggc	ttgtagccga	acgcgccctg	gccctcgcgg	accttgttga	10800

tgaccaccgc gccctcgacg ccggcggttg aggcgatctt ccgcagcggc tcctggagcg 10860
cgcggcggat gatctccacg ccgaagtcct gctcaccgcc cagcttcagc ttctccagcg 10920
cgggcaggca gcgcaggaag gccacgccgc cgccagggac gatgccctcc tcgacggccg 10980
cgcgagtcgc atgcagcgcg tcctccacgc gggccttctt ctcttcctc tcggtctcgg 11040
tcgcggcgcc cacgttgatg acggccacgc cgcccaccag cttggccagc cgctcctgga 11100
gcttctcgcg gtcgtagtcg ctggtgacgg tgcgatctg cgagcggatg agcttgatgc 11160
ggccctcgat ctcgctcttg gtgcggcgcg cgtcgacgat ggtggtggtg tccttgcca 11220
ccgtgatgcg cttggagcgg ccaggtcgt tgagcgtcag gttctcgtac ttgtggccca 11280
gctcctcgct gacgaccatg ccgccgtca ggggtggcgat gtccttcagc atctccttgc 11340
ggcggtcacc gaagcccggc gccttcaccg cggccacgtt cagcacgccg cggatcttgt 11400
tgaccaccag ggtggccagc gcctcgccct cgatgtcgtc ggcgatgatg agcagcggct 11460
tcccggagcg cgccacctgc tcaggatgg gaatcatgtc ctgcatcgac gagaccttct 11520
tctcgctgat gaggatgaag gggtcgtcca tgacgacctc catgcgctcg cggttcgtca 11580
cgaagtacgg agagacgtag ccgcggtcga actgcatgcc ctccaccacg tcgaggttgg 11640
tctccaggcc cttggcctcc tcgacggtga tgacgccctc cttgccacc ttctccatcg 11700
cgtccgcgat gatggagccg atggtctcat caccgttggc ggagatggtg ccacacctggg 11760
tgatggcctt cttgtccgcc gtgggcttgg acagcttctt cagctcctcc acgacgacct 11820
cgacggcctt gtcgatgccg cgcttgaggt ccatggggct gtggccggcg gccaccagct 11880
tgaggccctc ctcgtagatg gcacgcgcca gcaccgtggc ggtcgtggtg ccgtcgccgg 11940
ccttgtcgga ggtcttcgac gcgacctcct tcaccatctg cgcgcccattg ttctcgaact 12000
tgttgtcgag gtcgatctcc ttggcgacgg tgacgccgtc cttggtgatc gtgggcgaac 12060
caaagctctt ctcgatgacc acgttgccgc cttggggcc cagggtcacc gccacggcat 12120
ccgcgagggg ccggacaccg cgaaggatgg cctcacgcgc ggactggtgg aagaaaatct 12180
ccttcgctgc catcgcaacc tcctgaaatt actgggggat ttaactgtga cgacgctggc 12240
cgtttagcact cggccatggc gagcgccaac ataagggccg gcgaaatgat gtcaaccagg 12300
aagggcgtgc ctgtgggcga gcggacgggg ccaggggccc tcggggccgt ctggccgggg 12360
ggctactcgg cgaagcgcac gaggttgagc agctgaccca tgtccagggg tttgcgcagc 12420
gtcagcgcca cgcccttgcg caggccccgg tcggtgacgc cctcgtcctg ggcgcccgtc 12480
atcagcagca cggaatggc ctggaagcgg ggtcggcct tgagcatctc cgtgaagtgg 12540
atgccgtcca ggtggggcat gaggtaatcg gtgatgacca tgcccacctg gttgcgctcc 12600
aggatgtcga gcgctgcag ggcacccgcc gccagtaga ccgtgtagcc atgcatctcc 12660
aggaagcgag agagcatggt gcacagctcg agatcgtcat cgacgacaag gatgttcacg 12720

8/38

gggcgggccg tcacacatgg acgggcgccc gaacctaaca gccgtgcccg tcgttgacaa 12780
cgcacagcgt gccttcatat gtgccggccc atggggaagc ggaccgcgaa gcgagagggg 12840
gtgggtggcg tggaggctcg gctggatgcg gtggcggacg ctttcgaggc cggggacttc 12900
gaggccgcgc tggccggggc ggagggcctg ctggcggacg cggccgagct tcccagaggcc 12960
ctccacttcc gggcttcggc cctggtggaa ttggggcggc tggaggaggc cgggaaggcc 13020
ttcgggcagg ccctgaaggt ggcgcggag gacctgaaa tcctgctgag cgccgcggac 13080
tgccctggtgt gccgcgccg ggaggaccgg gaggcggtgg cggaggggct cgccctgtgc 13140
gcgcgaggcc ggcgcctggc ccagaaggcc gacgacgtgg agatgctgta cgagttcctc 13200
ctcctggagg gcatggggct caaccagatg ggcgagtgcg ccacggcgct ggtgagcctg 13260
gacgcggcgc tcggccacat gccgcgctcg ctggacgcgc aggtcgagcg gggcatcgcg 13320
ctgttcgagc tgtgccgctt cgacgaggcg aaggccgctt tcgagaaggt gctgaaggac 13380
gcgccggacg acccggtggc gcaccactac ctggggctca tcgccgagcg gcgcggggac 13440
gagaaggagg cgaagcggcg cttcgacaag gcgcggggcg tgggtgcccga ggagttcccc 13500
ccgccggtgg agctggccga ggcggagttc gaccgcgcgg tggaggacgc ggtgaagtcc 13560
ctgccccgcc acgcgaagca gtacctggac aacgtcacca tcgccgtgga ggacctgcc 13620
tcggacgagg atttgctggg gcaggaccgg ccgctgtccc cgagcatcct cggggtgttc 13680
cgcgccacgc ccgtgggcca gcggagcgtg atgaacgcgt atgagctccc cgctccatc 13740
gtgctctacc agcgcaacct ggagcgcttc gcgaggacgc gcgaggaaact catcgagcag 13800
attggaatca ccgtgatgca cgaggtcggt cacctcatgg ggctggatga ggacgacctg 13860
tggcagcggg ggctggactg acgagcgagg cccgcgtggc ttccagcgtc tggcggatgc 13920
cctcgctgta gggcgtcttg cgcacgctgc ccagcagctt gtgcagcgcc gagtcgtcca 13980
tgaggacggg ctcggtgagc aggtagtgca tctcgacat ctcgcgcatg aacgggttga 14040
agaggccgat gaggcggagc atccccctcc ccgccgtcat gaacttcgcg aggtggccgg 14100
cctgggcgta gatctcctcg acgagcttgc gctgtgaggt gacgcggcg ccggcgaggt 14160
tccaccagcg gccgtaggca ccgggggctg ccatgagggc cgtcacggtg ggaccacgt 14220
cgggcacgta gacgtactcg tgcggcgggt cgatggggcc gatgagctgg gcgcgcttgc 14280
cctgggcggc ggcgacgaag gcggagtgga ggaagctctt gtcgatgccg gggccgtaga 14340
agtccgggag gcggaggatg gtggcgcgga gctttccggc ggcgtccgcg gcgaggagga 14400
tgtcctcctg ctcccttgcgc atgcggccct tgaaggtgtg cggctcgcg ggggtgtcct 14460
cggtgacggg cgtggtgcgg gggcgccgt acgggtacac ggtgccgatg agcaccacgc 14520
gctcgacgcc ctcggcgatg gcggcgcca cggtgcggcg catgagctcc ggggtggagct 14580

ggaagcgcca gtagtccacg ccgaccatgt agatgagcgt gcggatgccc cgggcggcgg 14640
cgcggtatgt gtccggcgcg tcgggggtcc aggtggcgat ctccgccttc ggggtccgcgc 14700
cgaactgcga ctgaagcgag gcgcgcgagc ggccaaccac ccggtacgcc ctccccctgct 14760
cgcaagagc cttcacgatg ctgtgtccga tggggcccga cgctccaaac aacgccacct 14820
tgtccgtatc catggtccgc tccctgggtt ccgaacggcg agcgttttat gaacgtcgtt 14880
cattgtgaac ggtgttcatc tactgctggc cggtcaggac gtcaagccct acagtgggga 14940
tatggggatt tcggagcgga aggagcgcca gcgggcggag ctgcgcgagc agatcctcca 15000
ggtggcgaag gacatggtga cgcgagaggg ctttgggtgcc ctctcgatgc gcaagctggc 15060
ggacgcggtg gagtacgcgc cggcgacgct ctacctgcac ttcgagaacc gggacgcgat 15120
tgcccgggag ctgtgcatcc ggggggtcca ggacctgctg gaggcgttcg agccggcggc 15180
atccgtggag gagccggtgg agcggctgta ccggctgggc gaggcctacg tgaagtccgg 15240
gctggagcag ccggagacgt accggtcat cttcatggag gacccgaagc tgtcgacggc 15300
gctgttcagg gacgcgccgg aggaacggcg ggggccgcgg tccttcgggg tgcgtgtgaa 15360
ggtgttcgag gacctgaagg cggcggggag gattgccgag gacgccgagc cctcgaagct 15420
ggccgaggtg ctgtggggccg ggggtgacgg catcggtggc ctgcggctga cgtgtacggg 15480
cttcaagggc tcgccgccgg aggagctggc gcggctgttg gtgtccacga tgggtgaacgg 15540
attgccgggg ttgaagcccg tgaaggcggc gcggaagaca cggtgagccg ggcgttggtt 15600
gacggggccc gtgggggtgcg gcagtaatcg acggtggtag ggccctccga cctggtggcc 15660
ggcccgtct tcaaaaccgg tgaggggcgc tgagaggcgt ccctggtgag ttcgattctc 15720
atgctctacc gccatgtttt ccttgggggt ttctgacgcg gccctgatcg ggttgacgga 15780
cgccccaccc gcctgtacct ctctgaccca agttttaccc aagtgagggg ggtacgggag 15840
ccgcgcatga cgaagtggag cgggaagtgg atcggaggcc gcacgtacac ggccaccgac 15900
ggaagcacgc gctggatcat ccgcaagacc gtcgccgggg tcgcgtacaa cgtcacgctg 15960
gacgtgcgga gcgaggccga ggccctggct gagctggcgg ccttccgccg caaccgggcc 16020
gcgtaccgga ccgcgagcca ggagcgccaa gacgaggccc aacaggccca ggaggacgcc 16080
gcgcaggcgg tgttcttgga cgaggatcgt gcccggcggg tctccagca tctgaagtcc 16140
agcggcaggt ccgacgagta cctgaagagc acgcgcagct acctggccgc gtgggcaacg 16200
gcgctcgcgg acaaggacct gcgcgcggtg accggaccgg cgttgaagaa gatcctggcc 16260
acctgggaca ccggcaagaa gtggaggatc atcgtcttca agtccttcac ctcgttcctt 16320
caggactcgg gcgaactgga cccgatggtg aaccccgac gcttcttgaa ggtccccctt 16380
gcccggccga accacgaact caagggttac agcatcgagc acgtccagca gctctacgcg 16440
gccctcggct cgcaggccat ccgcgatgtg ctgtgccttc aggccaagac cgggatgcac 16500

10/38

ggcaccgagg tggaccggct cgcgagcggt gacggcaaga tcgcggtgct gaaggaccag 16560
ggcgagatcg ccgccaccgt cacgttcaag cacaagaccg ggaagccgtt caccttgctg 16620
ctcgatgtgc aggggctcgc cgcgcgccag cgccttcaag cccgggggtt agcgccggat 16680
cggcagacca tccgcgagggc cgtggagcgc gtgtgcgcgc gcacgagcct ggagttcgtc 16740
cacttcggcg cgatccggca cagcttcgca acgtggctgg tcgagcaggg cagcatctac 16800
aaccggaagg gcggtggcct gtccctggaa gccgtcgctc aggcgctcaa ccacagcagt 16860
acgcggacca cggccctcca ctacgtcagc gccacagtgc cgcgatgta cgtcgtcccc 16920
atccgcctgg agcattcgaa tgatccatcg gagttgcgca tagctcgggc acctgcgggg 16980
tgaggctcat gaaccagaga gttcggagat cagtacgagg aggttcggtg atgcgtcagg 17040
ccccaggccg gagccttcaa cgttcttatt gcgcggaacg ctccagggtga ttaccgcacg 17100
cctcccacaa aggagccatt cgatgcagat cctggccgta accgaacctc ctagtttcgg 17160
ggggaagtgc accatcacgc acgtggcggg caacgttgctc gggtaacctc acctgcgcgt 17220
gtcctcgatg ctcaacggca gtgcaggagt cgtaaccgat ctccatggcc gcgcctcat 17280
gcaggcgccc ttcaactttg ggggtgtggga tctagaaacg cgacggaggg tcgcatcgc 17340
gttcgtcgag gcggccttcc cggagactcg cgcgcgaatt gccggagcca agccgaccag 17400
agccccatc acggttgcgg agttcggagg cgtgctgcgc tcgctgctcc cttccacgcc 17460
cactcacccc gtcgggacta ttcccgctc cattccgatc gacgccatca acatcgttgc 17520
ggcggacctg taccggaacc ggacctgcga tagggagcac ttggaagagt tgacgaaggt 17580
gctcaacacg gagttggagg cgcagtaccg cacgcgcctg gagtggtcag gagacacgct 17640
gcacctgcgc gccaaagtaga ccatcgcccc ctccggcgcg gccagcaaag tctcttcctt 17700
gatcgaggag taggatgcgg cgaactacgc ctcgatcact gccgatcat tcttgtcacg 17760
atgtcggggg gttagagcgt gtcgatccgg ctgcggatgc caggacaaga agttctgcct 17820
cctctatccc gttgggcat gtcgaggggc accaaaacc tcgctcgtga gcttcggccg 17880
cttcgttacg gcgactttca acgaggctca aatcgtgagc ttgcggatgg caacacctgg 17940
gttcggatac gctcggctg cggttccctc ctgcttccca cgcgcctcct tgggtcggac 18000
tcgttcaggc tgcttggttt gttgaggccg attgcctccc ttccgagggg cgcctccgc 18060
gagcagtgcc tcgatgctcg ccaccgtgat catcacggac cgtccgatct tcggtgcccg 18120
ctccaaccgg ccttgctcca ggagcttgaa gacctgtgac cgcttgacgc cgagcattgc 18180
cgacgcgcgc tccagcatca cagccgaggc gttctgtgtg accgattgag cgcgaagcgc 18240
acgaacctcg cgcttgatct cggccagttc ctggaggatg gcgcattcca cgggtgcggg 18300
ctcgtggctc tcaatgggca cggcaggtcg aaagactgga atgggaatag gaatgaggtc 18360

caacggtggc tccggccgag caagcaagga gcctgtacgg atatagctga agttctcgat 18420
catggacttg tactcaaccg gtctccgtcg gaggccgtgg atggtgtcga aggtgtggat 18480
cgagccggtg gccaccagtc cgtccctatt ctgaatgtat cgaatctctc cgaggtcctt 18540
ccgggcgctc gccctgcccc cctccgtgcc ggtgaaggcg aacgggaggg gcttggcccc 18600
gcacctgagc cccgttggac gccatggatg gcccggaaagg gtcacagcac cggctacttc 18660
agccctttac tcgtgctgcg gcgtcatcgt cctgacttcc ccatagcctt ccatcccttc 18720
cgatccttcc acctctctct gcaaccaggg ggagaccgaa gcagacagca tggaaggcat 18780
cgccgcagtc ggaacgcata cgttccaatc acgggtgctc aaggtccatc cgagcgccag 18840
cggcgatgct ctccatcacc ccgtcgatct cggctctcca gcgtttcacc gtgaccgggt 18900
acttgattcc atcgctccgc ttccagcccg tgttctcttc accggccaac tccttcatgc 18960
gttcaccgaa ctctgtgagc gagaacgcca tccggtcacg gccctggtag ttcaagtcct 19020
cgacgtagtg gccgtacagc gccgtggcct tgatccactg gtcccgtgc gcgcaccggc 19080
gctcgaagaa atcttccacc gagtcttgg ctogaatctc atcctggcgg gcctggacct 19140
gcgccaattg gccgaggagc gggctctccg ccttctcgtc cacgctcgtc cacatggcca 19200
ggaagtcgat ctgctccagt gccgtccagt tcatctgggg cagacagtcc agttgataga 19260
agcgtcgcat tcccgtagga tcgacgacga tctccttcac ctgcttggtg gacgcaccga 19320
tgaagggtggc gttgttggtg ccggtctgga gcttctggac accgagcacc cggtagctga 19380
tccgctcggc ggtgatgcgg ttcttgagca cgtcagacgc taccgctct gccttcgcca 19440
tttctgogaa gaacatgccg tagcagcgag tcaaccggaa ctgctggcgc tcatcggtca 19500
ggaactggag atcggcgggg aacccaacga gatcgccaag aggctgtagc aggctcttga 19560
tggccgtagt cttcccgcgc cgtgttttgc cgacgaagat cggcatcagg tggctcttga 19620
ccggcttccc gaacaacttc cgtttcacct gccaaatgaa gtgaagaacg accacgagat 19680
ctaccgggtc aaggcgtccc gtggttgcct gaatgaactc ctcggcgtca cctcggcgtc 19740
cgtcctcgcg gaacttcaac ttctcacgat actgctcgac gatcccccg cgttggtccg 19800
cctccacag gcgcacggcg ttaccgagat ttgcacgct gatgcatgtg cgctcgttcg 19860
tggaactcgat ctcgatctgg gacaagaact cggacgcagg aaacatccgg cccccaatct 19920
ggaagccatc gctcgcgtac acgaatgtga cgcggttctt ctccatccag tgttgaacct 19980
atgtgcgcgg ctogaagcgg ttctcgacct tcggcaacgc ctgcaccaat cggtagatga 20040
cctcgggctt gcagaccttc ttctcttca aggtcccagc ccaacttctt aagatgccat 20100
gctcgggtgga gaaccaaacc gagcaccaga cgagagcgcg cttgatctcg atctccttca 20160
ccccaccgtg cgcctcatgc tgttccacga gccagtcaca gaggccgaa ggaagcgccg 20220
ggcgcgagaa ggccaagtac tgcacgatca ggaactgacg agcccatgct tccgtcttgt 20280

12/38

ggtgatgagc	ggcaagcaac	ttcttcatct	cgcgatgcac	ccaatcatcc	acaccggatt	20340
tctcgtgctg	tgggtcgtgg	tccacgtacg	gaatctgagg	gtcacatccg	cgcgcagaaa	20400
ggcccttcgc	gacccgcaca	acttcacgaa	cgacgttcat	gttcgtgccg	tccatgtccg	20460
gcgcatcgcc	gcagacgaac	accgtgcggc	cctgtgcaag	gtatgatgcg	gcctcgggggt	20520
gcggtaccca	ctcgtccgtc	aggtcatecc	agcttcgtcg	ccgggcctca	acatcatggg	20580
ccatgctgat	gccggggagc	cccagcgccg	gggtgccttc	ctgcgcgagc	agcgccgcct	20640
tcttctcgcc	ctcaatgagc	gccatgggga	cttcttcgtc	ttgaagcatc	gggcagagga	20700
tcgcggtgtc	gaagatctgg	ggcgcgtacc	cgaaaggagc	ctcgtacttc	ttcaccttcg	20760
ctggctcccc	ttcgacctca	tcctgggtgt	acgggcgcag	gatgtgggtc	gtgaagggtc	20820
cggttgggag	ggtgagctgc	gcctccttcg	gccggagcac	gcgcgggcga	tgccggttga	20880
ccttggcgta	gttgagcccc	gggtacacga	gcaccagccc	tccagcgtcc	cccttcatt	20940
cctcgggcca	gttcaacaac	ttccggttgt	tctcgtccgt	gtccgttcgc	caaccgagtt	21000
tccgcaccgt	cgagggcagg	aggcaggact	tcttcaggtc	ctcgcagtct	cgcgccgaga	21060
tcggaagctc	gttccactca	gcctcgttgc	ggctggattc	aacctcgttc	gccttcgagt	21120
cgggcagtgc	cttgctccgag	tcctcatgtc	ccattgggta	cacctctccc	atcaacgccg	21180
aagccgtgct	tcgacgcctt	gaaatgcagg	gtggctcttg	aacttgtatc	gattctggct	21240
acgtcgttgc	ggcgccctgg	agcggagccg	ggttcggtgt	ggctacaagg	aatatatcga	21300
atgtgtctgc	gcacctcgga	cggtgcgcc	cgagggccac	gggcatcaac	gccgctgcgc	21360
gacctcgcgc	cgtggacggc	caggggattc	gcgacgtaca	ccggacggca	accttcctgt	21420
tatgagccga	ggcaccgcgt	ggagcacgga	gtacctggac	gcgatgctcg	cgccctggga	21480
gaggtccggg	ctgacctcgg	agcagttcac	ggccacggcc	gcgccagggc	tgccctgcgc	21540
taccctacgg	ctgcaccgcc	agcgaaaaga	gaaggagatc	cacgcagagc	ttcagcggct	21600
tcgggatgaa	aacgcgacgc	tccggtcgat	gctcgaaagg	agcgcgttgg	tccttggtgg	21660
ttgccggggc	gcaaccgtca	gcgaccccaa	cggccacggc	attggagatc	gccatgtggc	21720
aaacggaacc	gaagcccctt	cgccgaccgc	gaccggcccc	ccagcattgg	agccggtaca	21780
gccgcgcgtg	gtgccgcagg	accaggttgc	cggaccctcc	cccgtgatc	acgatggctg	21840
tgtgctcgta	ggggaaatca	ccgagccggg	gccccgaag	agaccggcgt	tctcctggga	21900
ttgatggcct	cgcgcctgct	ggccgatgtc	ggacgccatc	taccgctggg	cttggtccgc	21960
gagcgccttc	aagagcttct	tgatctcggc	cgaggacatc	gtgcggatgc	ggtccacgag	22020
gcgttgccgc	tcgtcatcct	cgggaaacgt	caggagatcc	agcggcaaca	ctcccagggt	22080
gtccgccagc	accttcaacg	tgcgaaggcc	aagctcctgc	ggccagtgtc	tagaggtggg	22140

actcagggca cggagcagag gcagagcgcg agcaccaggc aggcgctccg ccgcgtcac- 22200
cctgccgagc cctgtggtcg cttgtaatac cctttgggtt tatgaatagt agacctcgga 22260
ccgcgagagc caccggcccc gtccccctca gccgaggccc caccagggag ccggaaggcc 22320
cgagccagtg cccgccgcgc gtcccttccc acctggaggc cgcacatgac cgatgacagc 22380
acgccgaccg acgaaacgca gatcgaggcc ccgacgccgg agaacacctc cgacctgaag 22440
ggcgccccgc cccgactccg gttcaccatc gagaccacca cgtacgagac cctccgcacg 22500
ctggagcggc ggggtaacgg cgaggtggtg ctcttgccg agcgtcacct ccccatgga 22560
ctcgccgggc tcgtcaccat caagcgccgt cgcaatcccg tcactttcga gcgctgccag 22620
cggctgatcg aggaggtcca gctctccttc cgcctccacc acccgccat cgcccaggtc 22680
catcacctga aaatccatgc cgaccggccg cagtcacg ccgagtacgt ggacggcccc 22740
acgctggaca ccatcatcag cctcgccacc atgcgcgaga agccgctctc cgcgcccttc 22800
gccctctaca tcgccgccga ggtggccgac gccctccacc acgcgcacac gctgagagat 22860
tcggagaacc ggccgctggg catcatccac cgcgacgtgg cgccccggaa catccgcgtg 22920
gcccgcagcg gcgaggtgaa ggtgacggac ttcggcgcta cctactcgct catggtgggc 22980
agggaggaga cgccgggcct cctgctcaag ggcgacgtgg cgtacgcctc gcccgagtac 23040
ctgaaccgta agcccatgga tggccggtcc gacatcttct cgctgggcct cgtcctcatg 23100
gagatgctga cgtgcaagca cctcttcgac gtggaggacg aaaaggcccc caacgccacc 23160
gtggacgtga agacggagga gacgccctcg gtgccctca cgcagatgat cgcgctcgtc 23220
aaccgctacc gccccgagga cgtggagaac gcgatggcgg gcctgccgga cgcgctcaag 23280
gccatcgctc acaaggccct ccagcgaaag ctctccgagc gctacgccac ggccgccgag 23340
atgcgcgatg cgctgcgagc ggcgctcgcg gcggagtcac agcccttcgg ccggaaggag 23400
gcgagcgagg agctggcgcg gatgctgtcg gaggcgctccg tcctgcgcga ccgggtggaa 23460
ctggacgaag aaggcatctt ccccaggggc ctggacgccg acgagccgac gcccgaccg 23520
aacgaagagt gagaacgggc ggcggtgcat tccttgaagc aagcgggccc cgggtgtggc 23580
ccgcgtcggt tcagcgcaac aacacctggg ccaccttcac gagggcatcg agctggtcct 23640
cgtccatctt gcgcgccagt ccggtgagct ggcgagctt gggcgccccg cctgccctc 23700
gtccccctt gcccgacttc gtccccctct ccgcgtccgc gatccccagc aactcgctccg 23760
aggagatgcg cagcaccgag cacatccgca gcagcgtctg gacgctgggc agcatcttcc 23820
cgcgctccag gcggctgtag accatgtgcg cgaggcccag cttctcggcc acctccgct 23880
gcgtgagccc gagctgcgtc cgggcctcac gggcggcact tccaatacgg gtcgccagct 23940
cttcattcat gcgtcggggg accaggacac cgagaaccag gtttggcgcg ggcaggaagg 24000
cccaagcctc ggggacgccc cttcctgccg catctcggcg gccatgtcga tccacgagca 24060

14/38

gggccccgctc	acgcggcgccc	cctattggccc	ggacccatcg	ggggcgggaa	cctccagggc	24120
caggccgccc	aggatgaagt	ccacggggac	acggaggaca	ccgcacaggc	ccaccagcgt	24180
ggggaggctg	ggcagcaacc	ggccacgctc	caggcggtcg	aacaccagcg	tcggcacgtg	24240
gatggcctcg	gccacctggg	cctgcgacag	gcccgcgcga	tgccggggcg	cgcggaacatt	24300
ggttccgata	tgctgggaga	gcttcgcgta	catgaggaga	cctggggagaa	agaggagaaat	24360
gccgcgtggg	cggcactccg	agcatacagc	tcaccagggtt	caacgatagt	atccgctcaa	24420
ggtttaccct	tctcccgcg	cacttcgcgg	ggggaaactcg	tacgagatcc	gctaccgggtt	24480
gagagcatgt	ccaggaacgc	cgaccccaca	cgccccgcgg	aaggagagga	cacttgagcc	24540
cgctctgca	tcccgctcgtc	cccccgggca	cggacatcgg	cgggtacctg	gtggaggaga	24600
agctgggagc	cgggggcttc	ggtgcggtgt	accgcgcccc	gcgtggaggg	cggctctacg	24660
cgctcaagct	catcccactc	tgggggctgg	ctgagtgggc	cgagcgcgaa	gtggccatcc	24720
tcctccggt	caagcactcc	aatctggtgc	gcatccgtgg	gcatggacag	tggccggatg	24780
aggcgctca	atccttcttc	atcgctcatg	actacgtgga	aggcgccgg	ttggacgtat	24840
gggccagaga	ggagaacccc	tcggcccag	aagtcgtgct	caaggtgcgt	ggcgtggcgc	24900
gcgggctggg	cgccgcgcac	cgggcgaagg	tggtgcatcg	agacctgaag	gagagcaacg	24960
tcatcgagcg	cacctccgac	ggagaggcag	tggtggtgga	cttcggcgcg	ggcgggtacg	25020
agagcgcccc	cagcatcacc	ggcggcgtgc	tgccaccggg	cacgccggag	taccgcgcgc	25080
ccgaggcctg	gcgcttccag	caggagcacg	gggacgagcg	tggccactcc	taccagcccc	25140
gcccctcgga	tgacctgtac	tcgctgggcg	tcgttctcta	ttggctcctg	acgggcaggc	25200
agcccttct	cccggacgag	gcccagggg	tggaggccgt	gctcaaccgc	gcccccaaac	25260
caccccatgt	gctcaatcca	cgcgtccccg	aggccctgag	tgccgtgtgc	atgcgcctgc	25320
tggccaagga	gcccgaggag	cgccacccgg	acgccgacgc	gctgtgcgcg	gagctggagg	25380
ccctgctggc	ccaggcggac	gagtcctggg	acgtgaaact	ctgcgacgcc	tatggcccgg	25440
acaccgccac	cacgctcgcg	gcggtgccgc	acgcgggcga	agacgagctg	gtgcaatggc	25500
tgaagaagcg	caaggccccg	cctcgtcggg	ggccacgtcc	gccacacggc	gagggcgcac	25560
acgagccgga	ctcaaacgcg	gtggccatcc	aggcagagct	gccaccccca	gttcctgctc	25620
gccccggccc	catgccctcg	ccgcccccg	cacaagccgg	tcacttcaag	gccctgcgtg	25680
tggcgacctg	gatgggtctg	ctggtggtct	tgagcgcggg	cgcgctggtg	atcgcgcgca	25740
ccttggcgcc	ttcccctgct	cctgcgcac	gtcacgccct	ggagcctcac	cccagtgatg	25800
ccgtggactc	accacctgca	tccccgagt	tcgacctcgc	gccctggggc	cccggccagg	25860
aagtggcggc	gccatggatg	gcaccggaag	ctcaggaagc	cgcagccgcg	ctcgatagcg	25920

cgcccacccct agcggccgctc gccctccccg cgacgatctc cgaggagaag gcttccgtga - 25980
agatgaagaa gaacaccggc gtattctccg agcccagacc gcagcgcttc cggggcgga 26040
ccgttgga caa ggcgctcggc ctccgtgtcg cgggctgctt agcgatggcc tgcctggcc 26100
cccaggctccg cccgacgcct ccgccagagg cgtgcccgcc aggtgctgtc gaggccatga 26160
gcaggcttcg aattgagtat gacgaggatg tggtagccac cttcttcgcg accgccgagg 26220
aggggggttag caggaacatc acggtgcgcg cgggcccgc aacggttcac ctggggagac 26280
cattgggaaa tctaccggcc cggaccacgc tctccggacg cctcatcatc ggagaggatc 26340
gcgtttacgg ccgactaact gaagcacgca cgcaaaggg tgaccgcttc cctgtatgca 26400
tgcagatcat tgagaacgac tctattgggt tcaagaagga gatcaacgac ggctccaaca 26460
gcgcctcggc ctatccctat tttgacgtga aggcgctccg ccgcttcgag tgaaccccga 26520
ccaggagcaa catggggctg aaccaagtat caatgcaggg ttccactctg ctacttcaag 26580
gtctccgcg atgccatgt caccatcgt tgttctcctg gtgatagcgc ttttctccg 26640
ggcagccgcc gcgcaatcgc gtccagcagc cacgggcttg ggggtacgcc ggattgagtt 26700
ctcggccgag aacttcgagg caatgacgac accggaggta cagataagcc cgggggtttc 26760
caccacgttc gagttcaact ccgccctcct tcaggagaag gtcgcggttg aaggcgccga 26820
ccgcttctcg ctcggtggaca ttggacgaag cactctccg ctggtgccgt cggagcaagt 26880
cctgcccggc gaacgcctgc gactgacggt gcggttcctg gacggagcag cgccggtagg 26940
catggcattc atcctcgtcg cgcaccccg ctaggcggag cgcttggtcg aggtgtggcg 27000
caaccagcgc acggtggagt cctatcaaca ggaatcgaaa gaagcccag cggaggccca 27060
aaggtgtcac gaagagaacg agcgctcgc cgccgagtat gcgggaccgg gcggcctcgc 27120
tgccctcctt gcgaaccgcg tgatcggcaa gtcgggcgtt tcggtcaagc ccctcgactt 27180
cgagaaggaa gtccgccagc gtcctggaga cgcgatcagg attcttcgtg catggagcta 27240
ccgctccgcc aaccgcgtgt ccgtgctcat ggatttgac taccagaag ccgcgcaacg 27300
ctggatcgcg caggagcat cgttggtgag caagacaggt gaaacgctga gcatgctccc 27360
tgtatggcag gaagcgcggc ttgccaatgc ccgatccgg cgcctcgtgg tggaagcaga 27420
ggctgagccg gacgcggccc aggacacgtt caccctcaaa ctgtgggaat cgaacgggct 27480
gcgcacgatc accctatccg gggtagacatt ccataatgt tgagaacatc ctactatca 27540
agctagcacc cacatagcat gaagtccgcg ccggtgcgtg tcatgggccc ttcaaacc 27600
tgaaaaccaa taactacatc cttacaccat ccaaaccaag cccatgtct acagccgagc 27660
aaacgaagaa aataatcaag aagctgcgag ccaccctaaa cgaaggaagc tcagaaaaga 27720
ttcaagcact catccagtcc atcagactgg agaatecatc tcagtggctg aatttgatcg 27780
ccctcgact gggcccata gatcctagga atccaacatt cttcgaaaaa gtatccgcaa 27840

16/38

taaacagatt	ggcagaagat	cttgccgag	aagagcagca	cacgctacgc	ggatacga	27900
gcatcgcaag	ggccatgcag	atcgtatcgc	aagagtcgga	caaatacaga	ttctggaaat	27960
cttcgcgctt	cacaatgatg	tggcgccctat	ttacgtacgt	tgaatcccaa	tacataattg	28020
ttagcaatac	gattcgaaag	aagctcaagc	gggccccacc	gggctcgcca	tttgtcatgg	28080
cgagcccttc	aatcgcggtt	ccaggaaaccg	atgggaccgc	aacgacaagt	gtcgatcttg	28140
tttttgaatc	taccgcccga	gccgcccgaac	taatactccg	catctccatg	cacaggatcg	28200
acacccaacc	cccaaatgaa	aagtttacc	cccagacccc	ttatgacgat	cctgactttg	28260
gcaggctcct	cgcttgcgca	cacgcgtggc	atcgattcag	cctgctatgg	aaccaaattcc	28320
agtataccga	ctggcgacct	gagcatgatc	cacgcggctt	catctggaca	ccaaaaaaca	28380
aggcgcaata	cctccaaagc	gaaatcgga	acattcgga	gcgcctat	ttgcagcaag	28440
caatgctgtc	tgatggactc	ggtcctgggg	cagaactgga	gagtagaaga	gaagtcctgg	28500
aacgcctcaa	gaggagcttc	aatctcccga	agccaggga	aacatgggac	cttaagattg	28560
atattaaagc	cctgcgagat	ctcctcggtc	ttgatcgata	tgccacggct	cgcaagatct	28620
acatgtcgcg	tcgattccta	accccagtaa	tgggagcgcc	agcagatcct	gaactggaga	28680
agtcacagaa	ggagttgaag	ttgaacttca	cgctgcatga	actcgccatg	gctctgctgc	28740
aaaggaatca	cgcaacattc	aagaccaagt	cagcaataca	caacattggt	tttgcagtca	28800
ccaagcagcg	cctatgcaag	gtaatcagcg	aggcgtcggg	catcgaattc	gaattctgct	28860
tttcccatat	caacaacatc	gtcttcgacg	cacagagaaa	aaacctagag	atctgggacc	28920
agccgctcgt	accagtagat	gaccataggg	tggtgttcgt	tccttcgatc	attgccaaatt	28980
cccatcctat	tcgcacaata	gagcatttag	cttcagctcg	ctcaagctcc	gcgtttgact	29040
cccggggaga	gtattatgag	cagatgattc	aatcgaggct	tgaggaatta	acacaagcag	29100
tcgctgtgcg	gggcatcaaa	tttctgctt	cagatgggcg	gcagatcgaa	tacgacttgg	29160
tgctttattg	gcaagggcac	ctatttatct	tcgagcttaa	gtgcctgaag	atgacaacaa	29220
ctcctgtgga	tgagttcggg	gcgagggagg	agatcgagcg	cgccgagcac	caattggtac	29280
ggcgtgtttc	tattgccgaa	actgactgga	aggagttcag	gagccaagca	caggagttt	29340
cactccccga	gactccgcca	gcgcgcagaa	acatccatgc	cataatactg	accaacatca	29400
tgaactttac	gggccgaaga	actggagaga	tcgtgacggc	agacgactcc	tgctgttta	29460
gataacttga	cagccctcat	gttcccgtat	ttgatgctgc	cacgggtgcc	gaaatctcaa	29520
gcattgcacg	tttgcgctt	cgcaacgagg	ggccgccatc	tgcttcgggg	ctttggcaat	29580
atztatcaag	ccctccgcag	ttggagtgga	tgcggaagca	aattcggcct	gaataccacc	29640
cactgccacc	gctggatgag	ggcaactatc	actggaagct	catggtcgca	ttcgatccgg	29700

ccgacgcccc gacgactaaa tagtggttggg ttggtctcgt tgcgaagggt agcattgacc 29760
tcggctggac gctggattca gcgtctacaa caacgatatt gcaggggtgag gcggagggtg 29820
cgtagccttc gtttagaacc cgtggagaac ccgacccatg acgctatacg aactgtttga 29880
agccgatgat gaccgcgatg tccccgccga ctttccttgg ctcttgatcg agatccctga 29940
ggagcaactc cgccatgcgg cgcttctcct gaatgagggt gagtggcgac cctcgcacat 30000
ttctggcatc tggatccggg tggatcccga gcgcccagct cagaagcagc agcggcacgt 30060
ccacgtcgcc gcgaagaagc acatcaagat cccacgaag caggcttcat ggaacagaga 30120
caccacccgc catgaccgca agacattcaa tgcgaagctg ggctcgcaag ggagctatca 30180
agacgtggcg aagacggcgc tgggcctgcc gcccgaggct gtcctcgaac acgttggtga 30240
aaatccggcc gaacaacaga tgctgctcac cgaatccact gatccccagt tgaaaactga 30300
gttctacagg tggttggagg ccagcccag gcggtctaac ttcgaagcgc tcgtcgagga 30360
cgcggttagg ctcaatctgc cgctcgacct ggaatcgtaa ggacgtggtt aaatgcggtg 30420
atacgagcgg tgatacgagc gcctaacaaa actggcatag ggacccggct tctcaaggac 30480
gcagctgcaa ccaaggcgat tttgtaggca ttctttgggg ccgagacag cgcgcgggat 30540
ctccggcacc cagttcacgg acgtcacctg ccccggtgca gccaatggcc cctgaccctt 30600
gccgtcccca cctcgtacc caagctgtac ccaagatgaa cggacgcggg tggacctcca 30660
tggacgtcc acccaccg ctccagaggg aaaccggccg accctcaatc cctccgggca 30720
cctcgcggtt cgggggagcg gggcttcaaa accggtgagg ggcgctgaga ggcgtccctg 30780
gtgagttcga ttctcatggg gcgacgtgaa cccggccgtg ctccatgcgc tccacaacgc 30840
ggcggaagat gaggtcgata atggcgaggg actgagcctt caggagtggc cggaacagac 30900
cagggccacg tcagaagcga acgacctccc tttcgtaacc gaataacgaa acagaggcga 30960
ggatcacaac tttcgaaatc cgttgacgcc cgtgggcttc atttcaggaa tctgggaaa 31020
tgctctcct tgetcttctt gggcttctcc atggtagcgc ttcgcgcacc caggctctgg 31080
tactatccat catatggagc gctccctcgg aacgacagag caggctctct ataaggtacg 31140
gcagaagcgc cccttcaaca tcgtcattac gggcaagctg cgaggacgcg tggaggagca 31200
ggctgtgcgc caggggttgg tggccctgca acggcgacac ccgctgctcc gagcgaagat 31260
cgtcgcgggt actccgccac gcttcacctc gcaagggtgt ccccccattg cgctggaagt 31320
ggtggaacgc aaggcagagg agtccctggca acagcatgtc gaggcagagc ttgacaggcc 31380
tttcaactcg gacctgggc cgctcgctcc cttcatcctc gtgcggggcg agcaccgctc 31440
cgagctgctg tgcagctatg atcatctgat cggagacgcg cactccggca tcttcgccct 31500
gcgcgatctg ctgcaggta tggcctccca ggaccagcac ctgcccagc tcgcccgaag 31560
gcccgcctat gaggagctga ttgggcccgt ggtaccgggc acccgactgc ttggcgccgc 31620

18/38

cgTgcgcgg	ggctcctcgg	cgctgctgcg	cttgtctccg	accctcaata	cgctcacaga	31680
gcggtctcgtc	gcaccgcggg	gcatggctcg	cagcagcgcc	gagggacagg	cgctctacag	31740
ccaccgaatt	ctggagcccg	agcagctcgc	gcggtctgctt	gcgcgctgtc	gagagcaagg	31800
gtcctccgtc	cacgcagcgc	tcggggccgc	gctgctgata	gctcggggccg	agtcgcaggg	31860
tgccaagaag	cgggtgacgc	tcaccctcac	ctctgcgttg	gatgctcggg	agcgcttttg	31920
cgtgggtgag	gacttcggcc	tggtcaccac	ggggaagacg	gacttcttcc	gcgcgcgcgg	31980
cagcacccca	ttctgggagc	tggcgcacag	actccgatcg	cctctgcagg	cagcacggaa	32040
gaagcggagc	cacttgaggc	tggtccgcct	gactctggag	atatccgcgc	tcaccgtgga	32100
catctctcc	cagccctgga	tggaacgcgc	gacacggctg	ggcctgcact	ccatgctcgc	32160
cctgagcaac	gtggggcggtg	tggagattgc	cgcccgctat	gggaacctga	cgctcgaggg	32220
ccttggtctc	acggggacga	ccgcggccca	gttcgacgtg	gtgctgacgg	cgatcacctt	32280
tgcccagcgg	ttggaggcca	gctttctctt	caacaccgcg	cacatgcagc	gcgagcaggt	32340
ggagcagctc	gcctcgcgca	cctgggagct	gctggccgag	gcgacgcgct	gagcggctac	32400
tccggctgga	tggtgtacat	ccgcttgacc	tcttcacgcg	gcagctccat	gaccttccag	32460
ggatcccagt	ctccgaagag	gtccacgctc	atcgcgctgc	ctcgtccac	cgcacgccgc	32520
agctggggaa	tgggaagcgc	ctcggcgagc	aggcgaacgg	gcaccaagga	cttgactagg	32580
gccggatcat	agttgcgggt	gcgggggccc	aacatgaagc	ccaggctgaa	ggccgtggtc	32640
agcatttcct	cccagatgga	gatgccatag	ccggagagca	ggtggacgaa	gtcatgcacg	32700
gccgagtgca	cgggcagggc	gtgagggtga	cccggcagcg	gcacgccatg	ggagtggcag	32760
tacttccaga	actcgcggcc	aaaggaaccc	tcgggcaggt	gctcgagctg	ctgaaagcgt	32820
tgcttgagg	ctggcgctc	attgcgctcg	cgcagcatgt	catagagctt	ccagggcagc	32880
tcgcccttcg	gcagcccca	catgcggcag	aagtactcaa	tataggcagg	cacgaggaag	32940
gcggtcagtc	caacggacag	ccacagcttg	gagcgcgcgt	ccgcggccag	ggaacgctcg	33000
cgcggtccag	gggctcgctg	gtacttccgc	agaaatacat	ccgtgagtat	tctattcga	33060
acaccttggc	tcgttatacg	gctggggcag	ctttcaggca	ggtcgctgtc	tcctggagaa	33120
agccatacgg	atccttgggg	tcctcggcca	tggactccac	ctggagcagg	cggcctaccc	33180
ccagccgctc	ggcgtgcgcc	aggatgactt	ccagcaccac	cttgtcctcc	aaggaaaacc	33240
cgggtggagtc	gaagacgggtg	aggccatcgc	ggtgagagcg	ccactgctcg	ggagagcggg	33300
tcagttcaaa	gaggctgggg	ccaatctctc	caggctcgag	ctgctggcac	tccccctcca	33360
ccacggcctg	ggggagatag	tccgggcata	ccagggcgcg	acgcagcatc	tccaggggca	33420
gctctgtctt	gccgggcaga	tctgagccca	cggcgttgat	atggagggtg	gccttgatca	33480

gacgcccgga caccaccggc cccgcgccc cggcgacgga ggtggccgtg cagaggatgt 33540
ctgcctcctg ctccacgcgc gtcaggggga ccgccaccac ctccaggccc aggaaggcca 33600
cacgccgggc aaagcttcgc atggcttcgg gagcaatgtc gtgcaccagc acccgtecca 33660
gaggaaagac gcgcgagagc gcatgtgcct gggtgacggc ctgagcgcca caccctacca 33720
ggcccaatac ccggctgtcc tcccgcgcga gcagccgact cgccacggcg gaggccgccc 33780
cgggtgcgtag cgcggtagga aaggtgccat cacaaacggc ccgcaggtgc ccggtcgtca 33840
catcatacag actgagggag gaaagaatgg tgggcaagcc cagctcctcc gggctgcggg 33900
ggttgtagcc caccagcttg atcacggcgt ggcgcccac ctgcatgacg ggcattccact 33960
ccaacacgcc ggggtggggg tgctgatact ggaagccctc gcgcttgccg agctcgggtgt 34020
tggcggggtc gaagtgtccc agctcgtact cgaggggtga gatgacctca tccatcagcg 34080
cgtgcagccc cgttgacagg acgatctggc gcagttcctg ctcaccagg atccttgtca 34140
tccgatggc tgaccccgcc cctttcagct gcatgacagg cctcactgca ttcacggctc 34200
cagttcaggg ccgggcaaaa tctatcttca acatcccctt gtccaggacc gcgtccacgg 34260
catccagctg cgccaccgag cgcggcagca tgacgtggcg gcggaagccg cccacgcgga 34320
tgatgaggtc attgccctgg cgatccagga caatgtctcc gcggtcggca cctggcatga 34380
ggagctgcag ccgatagcgg ccctcgtgct gcttggcgaa ctgataggc tgcccggaga 34440
tgtagcgctc ggtggggctc gagtcccctg agagccgctg cgccagatcc tccaaccgct 34500
ccagccccac cacctcgtgc tccagcagcg ggagcctggc caccggcatg ggggagaagt 34560
actcctggat gtgatccagg taggcagatt gagacttggt ccactgagca aagtattccg 34620
tgtccggcag cagacgattg accacgacct ggtccaccgt catgccatac atgttgaagt 34680
agaggtaggc ccgctgtgtc tcgcggatga ccattttttc cgcgctggag accaatcgca 34740
ccgtggtgcg attgccatcc gtgagcaacc cttcaatgcc cttgacgcga tcgaagagct 34800
gctcgaggga ggcaaagtag ctgtcctccg gcagatcata gctcgtgagc cgactggcca 34860
ggggccgggc cagcttgacg agcgtgcgat ccacattgaa gcgcccggcg atataccact 34920
gcagcacgct ggtgatgttg acgaagcgca gcgactcgcc cgtgggcgga cagtccacga 34980
tgatcacgtc atagcgcccc tgctgcacgt actggttcag gtagatgagc gagatgacgt 35040
cctccgtgcc cgggagaatg gcgacctcct cggccaccat gttggatacc ccggtggagg 35100
tcagcagcac ggacatgtag ttatagactt cactccactg ccgctggagt tcgtgctgga 35160
tgtcgatctc ttgcagttca aggttggggg cgaccttctg gggcagtcct tcgttgaagt 35220
cgaagagctt gcgatcgaga tcgaagctgt ccgagaggct gtgggcaatg tcgaaggaca 35280
tcaccaaggt gcggtagccg cgcttggcag cggccaccgc ggtggccgcg gatgccgtcg 35340
tcttgccctac cccacccttg ccagagaaaa gaagaattct tgtcatggat gccccgcgtg 35400

20/38

cctgccggca	acgggccgcc	atggcgctct	ccccctccat	gggtcacctg	ttgcgtctgt	35460
ggaactcatc	actatcagcg	acatgccaga	aggagcaagc	tggacggtgg	cttgacttgg	35520
tatgacgctg	gacagtaggt	atgccaatag	agggtggggg	ctttatgaaa	tcgatggatt	35580
ggctatcggg	cacagccccc	ccccttgttc	aattccagga	ggggtgcaat	tgacagccac	35640
agcgactgcc	gagcacgcgc	tgaaagccct	ccgggccaaag	ggactcgcca	cgtcggacga	35700
ggcgctgccc	gcggtgcggc	acgccagcgg	cgaaaaggct	cccctctcat	cctcacagcg	35760
gcgcatgtgg	ttcctggagc	agctcgagcc	aggcaacgcg	gcccagcatt	tgttgcaaag	35820
tcacacata	caggggcctt	tgcaggtaga	ggggttgccg	cgggcggttg	ggctcatggt	35880
caagcgccat	gaggcgctcc	ggctggtggt	tgccattggc	gcggcgggcc	ctgagcagcg	35940
tgtgcggccc	gagtgggtgg	ccgagctgcc	tctcagcgat	ctgagcgggg	tcgcccaggc	36000
ggacagggct	cgggccctgg	cggagctggc	gcgcgtggag	gctgcggcgc	ccttcaacct	36060
gcagcaaggc	ccactgttcc	gagtccggct	ggcgcgactg	gcggaaacgg	agcatgtgct	36120
gctggtcacc	ctgcatcacc	tgatctcgga	tggagcctgg	agctgcgagg	tgctcatcaa	36180
agagctggcg	atgctctacg	gccagcacgt	ggaaggctcg	agcccggagc	tggcgccact	36240
gcccgtccaa	tacagggact	acgctggctg	ggaggcaagc	ttcgcgcccc	cgggagaggc	36300
cttgaggagg	tggtggcggc	agcggctcgc	cggggtgccc	acggtgctgg	agctgcccac	36360
ggaggggagc	cgccccctca	gacagacgta	ccgggcaggg	cggggtggcca	tcaccgtgca	36420
gccgcgcctg	cgccaggcgc	tggaggacct	ggcgcaacaag	gagggcgtgt	cactctttgc	36480
cctgctgctc	accgccttca	ccacgctgct	acaccgctac	tctcgccagg	aggaattggt	36540
ggtgggggtg	cccgcctcgc	agcgcccccg	gcccagctg	catggcctca	ttggatactt	36600
cggcaccccc	gtggcattgc	gctcgcggtc	ggagcctcac	acccgggtgc	gggaggcgct	36660
gcggcagctc	gatcaggagg	tgcgtgaggc	caacgcccac	gcggcgctcc	ccttcgagcg	36720
gctggtgagc	ctgctagaca	ttgcccgcag	ccccagccgc	caccgctgtg	tccagggtgct	36780
gttcgacctg	ctgccggagc	agccccagcc	taccgctggc	ggctgcgcct	tccggccctg	36840
ggagtccttc	accggcctgg	tggcgatatga	tctcactctg	ctgctggagc	cgcgcggtga	36900
ggggctggag	ggggcgctgg	actacagcgc	ggacctgttc	agcgaggcgc	gcatgcagcg	36960
cgccgccact	cagtacctgc	acctgctgga	gcagctcgtg	gagcggcctc	acgagaggct	37020
ctcgcggtg	gcgctcctca	cgcctggtga	gcacgaggct	ttgctggcgg	ggcacggcct	37080
ggggcgggccc	gtggagggcg	cgcagggtgct	ggcccaccac	cgcttcgagc	accagggtgcg	37140
gctcaccccc	catcaccccc	cgttggtgctt	cggcccgcag	gtgctctcct	acgagcagct	37200
caaccgcgcg	gccaaaccgc	tcgcccaccg	gctgcggcgc	ctgggcgcg	gcccggacac	37260

cctggtaggg ctgtgcgtgg agcgetccct ggagctgccg gtggcactgc tggccatatg 37320
gaaggccggg gccggcttcc ttccgctgga cgtcaatcag ccccgcgagc ggctcgctt 37380
cctcctggga gatgcgagct gccgcatacct cctcaccag gagcacctgc tccagcgct 37440
gccccccacg aacgccgcgc tectgtgtct ggagagggag gcggaggcac tggagcgcg 37500
gccccaggag gatgcgccac acgaggccgg gctcgacaac ctggcctacg tcatccacac 37560
ctctggctcc acgggcacgc ccaagggcat cgcgatggtg catcgctgtc tggccaacct 37620
ggtggcctgg cagctcacc atgagcggt gggtggccc tcacgcacgc tccagttcgc 37680
ctcgctcaac ttcgacatct gctaccagga actcttcacc acctgggccc ctgggggcac 37740
cgtggtcatg gtgacggaag aggtgcgcag ggatcctgcc cggctgctgg aagtgctgga 37800
gcaggagcag gtgagccggc tgtacctgcc cttcatcgcc ctgcagcagc tcgcgcgggt 37860
ggcggatgag cgcggcgcg ccctcgcca cctgcgccag ctcatcacc cgggggagca 37920
actccaggcg acccccgagt tgcagcggt gctctcgcg atgccggagt gcacctcca 37980
caaccagtac ggcccctcc agtgccacgt ggtgacgagc catgacctga cagcgagcc 38040
gagccgttgg ccaaggctgc cgcgcgtagg ccggccactg gccacctcc gggtgctgtt 38100
gctggacggg gagcagcagc tcgtgcctcc gggtgtggcc ggagaggtgt tectgggcgg 38160
ccccgcctg gcgcgcggct acctgggacg tccggagcag acggctgatc gcttcgtgcc 38220
cgatccattc tccgcgagc ctggagccc gctctaccgc accggggacc tggcgcgcct 38280
ccgagaggac ggagcgctgg agttcctgca acggatggac gcgcaggtga agatccgggg 38340
ctaccgcac gagcctgggg agatcgaggt ggtgctgtgc gagcacccc cgtgcacca 38400
agcgcacgtg cggccctacg tggacagcgc cggagagcgg cgcctggtgg cctacgtggc 38460
cgcacggctc gaggacacgg acggggcgga gacagagcac gtggagcgct ggcgcgcggt 38520
gtgggatgag acctatggcg gcccaagcgg cacggccttt gacctggccg ggtggaatga 38580
cagcgtgcgc ggcgagccgc tgcctccgga gcagatgcga gagtgggtgg agacgacagt 38640
ggagcggctc atggagctcg tccccggcg ggtgctggag ctgggctgtg gctccgggct 38700
cctgctccgc cggctcgccc ccgctgcga gtcctactgg ggacagcagc tctccccggt 38760
agccgtcgag cggctgaggg agcaactcca gacgggcggc tccccgctcg cgcagcgct 38820
gcggctcatg gccagccgg cggtgactt ctccgggctt ccggaagccg gcttcgacac 38880
cgtcatctc aactcgggtga cgcagctctt cccagcgtg gactacctgc tgcgggtcgt 38940
ggagggcgcg ctgcgcgtgc tccaaccggg cggaacctg ttcatggcg acgtgcagaa 39000
tctgcggctc ttcgagctgt tccatgcctc ggtggccctg gagcaggcgt cagcggacct 39060
ggagggcgca gcgtgctcg ccgcacccg ccagcgcag ctgctggacg aacggttgta 39120
cgtggaccg gacttcttcg ccgcgctgc cacgcacttt cccagctgg gcgcggtccg 39180

22/38

gctgcacctc aagcgcggga gcgggaggaa cgagatgaac cgcttccgct acgacgtgga 39240
gctgcagctc gccccgtag cgaaagcagg ccccgcgcaa gagctgccag agctggactg 39300
gaggcatgag gggctcggcc tggaaaggct ggagcgcagc ctggcggagc gaccggcggg 39360
gctggtgctg cgcaatgtgg ccaacgcccg cacggcggac gaggcggcac ggctggcgct 39420
gctgcggact ggcaacagcg tgggcccggct gcgggcactg ccggcgggtga catcctggaa 39480
tccagagcag ctctggaggc tggcggaggc agccgactac acgtgccacg tcacctggag 39540
cgcccaggac gaggaaggcc gtttcgacgc gctgctcatg gccccgcggg ctggctcctc 39600
gcgcccggcg gcctggctga cgcctcccc gcctccccct cgaccctgga agagctacgc 39660
caaccagccc ctggcggcca gccgcggcg caccctcgtg ggagtgtgc gaagccacct 39720
cgagcacaag ctgcccagat acatggtgcc ctcttcttct gtgctgtgg acgcgtgcc 39780
gctcaagccc accggcaagt tggaggtggc ggcactgcct ccccccagac ccgcacaagc 39840
cgagcaaacg ccaggccacc tcgcgcgcg gacccccacc gagcgcggc ttgcccagct 39900
gtggcggcgc gtgctcggcg tgcctcgcgt gggggtggag gacaacttct tccagctggg 39960
gggcccactc cttctggcca cacggctctt gtcgtcatc cgcggagagc tggggctgga 40020
gctgcccctg cgggtgctct tcgagcagcc caccctggcg gccatggcgg gttgtctcga 40080
ggctcagagc tggagcacc aggtctccca cctcctctt gcctccatcg aggagggaga 40140
actttgagcg ccggagcgtt gctcgccac gccgcgagcc tgggtgtacg cctgtgggta 40200
gagggcgagc gactgcgctt ccaggctccc ccagggggtga tgacgcccga gctgcagtcc 40260
cgcttgggag gcgctcgcca tgagctcatc gcactgtac ggcagctcca gccctcctct 40320
cagggaggca gcctcctggc ccagtggtc cgaaacgggc ggctggcgct ctcttcgcc 40380
cagcagcggc tgtggttcca ggagcagctc caccctgagg ccccgccaa caacctcacg 40440
ggggccgtgg tgttcaccgg cccgctgcac gtggccgcgc tgctcggggc ggtggctgcg 40500
ctggtgaggc gacacgaagc gctgcgcacc accctgggcg aggagggcgg agtcccctat 40560
tcgctcatcg gcgagccgtg gcagccggcg ctggaggtg aagcgctacc gggtgccacc 40620
gtgggcgagc ggctggagca ggcccagag gtggccctgg ccgagtcgcc gcgccgttc 40680
gccctggaga cggagcccca cctgcgcgtg cgctgtctgc gcctggccga gcagcagcac 40740
gtgctggtgc tcagcctcca ccacatcgcg gcggatggag tcgggctgca ggtgctcgaa 40800
caggagctgg cggcgtcta cggcgcgtg agcgtgggg ccgagccgcg gctacctccc 40860
ctgcccctgc aggtggcgga ccttgccgac tggcagcgcc ggtgggtgga gggagaggag 40920
taccaggtcc agctcgctta ctggcggcga cagctcgccg ggctcacgcc cctggaggta 40980
ccaggagatc atcctcgccc gcgtatcccc tccatgagag gcgcggaggt acgggcgccc 41040

ctgctctccg ctccgcaagc ccagggtgctg cgcgcgctgg ggcaaggcga gggcgccacc 41100
ctctacatga cgctcctggc cgccctgggc gtgctgctgc aacgggtggac gggccagcac 41160
gatatggcgg tgggcagcgc cgccgccaac cgcaaccgcc cgggcttggga gggcatcctc 41220
ggcttcctcc tcaacatcgt cctgctgcgg ctcgacttga gaggccgcc ccgcttccgc 41280
gagttgctgc gtcaggcacg cagggtgtgt gtggaggcct atgcccataca ggagttgccc 41340
ttcgagcacc tgggtggaggc gctccagcca ggcagcgagc gcggggacag ctcgctctat 41400
cgcgtggcgc tcgcggtgag cgacacgcca tggatgcccg gccacgggct gaagctggag 41460
ggggtgcaag cgcagccgct cgactttccc cggggggtgc tggacctgga tctccacctg 41520
tgggtgtatg acaccgggga ggggctcacc gggcggttgg agtacgccgt ggatctgtat 41580
gaggagccca cggcaaggcg gctcctggag ggcttccgcc aggtgctgga ggcggtggtg 41640
gaggcgccgg atcggcccggt gccggagctg ccggtgctgg gagagcagga gcgccaccag 41700
gtgctctccg ggtggaaccg cacacagcgc ccctatccgc gagaggccag cgttcacggc 41760
ctctttcagc agcgtgccct ccaggccccc cgggccgtgg cagtgggtgta tggcgagcgc 41820
agcctcacct acggcgagtt ggcggagcgg gcccgccggc tcgcacaggg tctggtggcc 41880
cgaggcgctg ggcgaggaga cctggtcgcc cttcggctgg agcgtcgcc ggagcagggtg 41940
gagtcctatg tggccgtact ccaagccggg gccgcctacg tgccgctgga tcttccctac 42000
cccgtacagc gccaggagtt catgctccag gacagcgggc cgcggttget cgtgcacagc 42060
ggcccgtgct ccttcgctcc ccagggtgct gccacgctgg atctccaggc atggcacccc 42120
gctccctccg acggtggcga gccctgccc cagtgcagcg gagaagatct ggcgtatgtc 42180
atctatacgt ccggtccac cgccagccc aaggcgctgg ccgtgtgcca ccgggccatg 42240
accgcctgg tgtgcaacac ggactacgtt cagctcgggc cagaggatcg ggtcgcccag 42300
gcttccaatg cctcctttga cgcggccacc ttcgaggtct ggggcgcgct gctcaatggg 42360
gcacggctgg tggggctcgc cacggaggag gccattcagg cgcgccggct ggccgaggtc 42420
ctgcgcgagc agcggatctc cgtgctcttc gtcaccaccg ccctcttcaa ccacgtggcc 42480
cgggagcaac ccaggcctt cagcacctg cgtacctgc tctttggcgg cgaggcagtg 42540
gatgcctcca gtgtgaggcg ggtgctgaaa caggcgctc caggacacct tctacacgtg 42600
tacggcccca cggagaacac caccttctcc accgcctggc ggggtggaaca cctggccgag 42660
caggcgcata ccgtgcccac gggcatccc atcgccaact cgcggctgca cgtgctggac 42720
gaagcgctgc agccggtgcc ggtgggcgcc atgggcgagg tatacctggg aggtgatggg 42780
ctggcgctgg gctactggcg ccaccccgag gccaccgcgg agcgttctgt gccggatccg 42840
cacggcctgg agcccgagg ccggctctac cgcacgggag atctggcgcg gcggcaggcg 42900
gatggggcgg tgggtgtcgc tggccgcgtg gacaggcagg tcaagctgcg aggccttccgc 42960

24/38

gtggagccgg	cggagatcga	gtcgcacctg	tgcgagcact	cggaggtgtc	cgcggcgggtg	43020
gtggagctgc	ggggcgaagg	ggccctgcgg	cggctggtcg	cctatgtcgt	ccccgcgcc	43080
ggtggccgcc	caggtgcccga	ggagctgcgc	accttctctcc	gcacccgact	gcccaggtac	43140
atgctgcccc	ccagcttctc	cctgctggag	gcactgcccc	tcacccccaa	cggcaaggtg	43200
gatcgctcgg	cgctacccga	gagcttcgag	gaggcctccc	gcgagcaggc	tccggtggta	43260
cctccacgag	gccccgtgga	ggcgctgctg	gtggacatct	ggagggaggt	gctcggcacc	43320
caacgggtat	ccgtgcacga	cgacttcttc	gacctgggcg	gacactccct	gctggccacg	43380
cgggttgtct	cccgtttgcg	agaggcgctg	caggtggagt	tgccgctcgc	cacctcttct	43440
gaggccccct	agctctcggc	gctggcggcc	caggtggagg	tctgtctcgg	ccaccggcag	43500
ctccgcccc	ctccgctcgt	gcccgtgtga	cgccaccag	agctgcccct	ctccttcgcc	43560
cagcagcggc	tgtggttctt	ccagcagctc	gcgccacaaa	gcaccgccta	ccagatcctg	43620
gatgcgtggc	atgtgcgggg	ccgggtggac	gtgggcgccc	tggagcgcgc	cctggagcag	43680
ctcgtgcgcc	ggcatgaggc	gctgcgcacc	actttcgagc	ctggcggcga	cggcgtgcct	43740
cggcagcgca	tccacgcgcc	ggctccggtc	ccgctgcggc	aggtggacct	gcgctcccat	43800
ggcatcgccg	cgcgggagga	ggccttgccg	tggatgcgcg	agcaggcgct	ccggccactg	43860
gagctggaca	aggggcccgt	gctgcgcgtg	tccctgctgc	gcttgaaga	ggcacagtcc	43920
ctgctcttcc	tcgagctgca	ccatatcgtg	ggcgatggct	ggtcgctgag	cgtttggagc	43980
cgagagctgt	cgcacctgta	cgaggcggcg	ctccatggcg	ctgagccggg	gttggcgccg	44040
ctgcccgtgc	agtacgcgga	cttcgcgctg	tggcagcgcg	gctggttgca	agggccgggtg	44100
ctgcgggagg	agctgacctg	gtggcgggag	cggctcgcgc	gcctggcccc	gctgcgcctg	44160
cccgcggacc	atgcccgcgc	ggaagtgcag	cgttcaacg	gcgcgacgta	ccgcttcacg	44220
ctcccgggcg	tacgggtgca	ggcgctgcgg	cggctcgggc	atgagcacgg	ggccacactc	44280
ttcatggtgt	tgctgcgcgg	cttcaatgcg	ctgctggcgc	gctacaccgg	gcagacggac	44340
atcgccattg	gcgcgcccac	cgccaaccgc	acacgtgggg	aggtggaggg	cctcatcggc	44400
ttcttcgtca	acaccctggt	gctgcgcacc	cgctggaag	gcaaccctc	cttcctcgag	44460
ctggtgaggc	gggtgaggga	gacaacgcta	gaggcctacg	cccaccagga	gctgcctttc	44520
gagaggctcg	tcgaggagtt	gcagcccag	cgccaggcca	accagaaccc	gctggtgcag	44580
gtgctgctcg	ccctgcagaa	cgcgccccgc	gagccgttgc	ggctggcggg	cctggaggcc	44640
gagcacctgg	agtacctggt	ggccaccacg	cgttcgacc	tcgagctgca	cctgtgggag	44700
gaggaagagg	ggctgagctg	catcgccgtg	tatgaccgcg	cgctgtacgg	cgcggggacg	44760
gtggagcggc	tggtgggcgc	ctggtgcacg	ctgctggagg	gcgtggccga	gctcccggcg	44820

cgccgcgtgg cggagctgcc cctggtgcct gcccgggagc tgcgacaggt tccgccgccc 44880
tccccgcgcg cggcgatcga aacaagcatc ggcgcccgt tctcggaggt ggcgcggcgg 44940
cagccaggcg ccacggccgt cacccaaggc gggcggcacc tcacctatgc agagctggag 45000
gagcgctcgg aacgcctggc ccgctacctg gcctggctgg gcgtgcgcgc gggcgacaga 45060
gtgggggttg ccaccgagcg cacgctggag cggatcatca gcctgctagg catcctcaag 45120
gcgggagccg cctacgtacc gctggatgtg cgcagccag cgcgtcggct gagectgctg 45180
gtgcaggccg cgggcgtgcg caccgtcatc gccgaggagc aggtcgcac ggtgctgtca 45240
ggcctggggc agccactgac cctggtcgat gccgcccagg agccggcctc cgcgcagcag 45300
gtccccgcgc tcgggcccga gcgctcgctg ggcggggaca tgctcgcta cgtgcttttc 45360
acctcgggct ccaccggcga gcccaagggc gtctgcattc cccaccgtgc cgtgctgctg 45420
ctgatccatg agccctccta cgtccagctc tcgccgcggg aggtcatgct gcactacgca 45480
ccgctggagt tcgatgcctc caccttcgag gtctgggggg cgtgctcaa cggggcaagg 45540
ctcgtgctgg tccctccgga acagcagtcg ctcgagagcc tgggacagga gctgagcacc 45600
cagggcgctca ccgtgctgtg gtcaccgcc gggctcttcc ggctcatggt ggaggagcag 45660
ctcaagagcc tgcgcggcgt gcgccagctg ctcgccggag gggacgtgct gccatgccg 45720
caggtacgcc ggctgcgcga ggccctcccc gagtgccagc tcatcaacgg ctacggcccc 45780
accgagagct gcaccttcac ctgctgtcac cgcgtgggca gccccagga gctgggcggc 45840
tccgtgccca tcggcacccc catcgacctg ggctgggtga gcgtgggtga cgagcggctc 45900
cagcccgtgc cggacggagc cccgggcgag ctgctcgtgg gcggccccgg gctggcttgg 45960
gggtacctgc aacacccgga gtcaccgcc gagcgcttca tccccgatcc gtcagccgg 46020
acgccggggg ctgcgtcta tcggacggga gacctggtgc ggcgcaggga ggatgggacg 46080
ctggagtctc tgggcccgtg ggaccaccag ctcaaggtac gcggcttcgg catcgagccg 46140
ggcgaggtgg aggccgcggt gtcacccac ccggcgggtc agtccgcggt ggtggtgggg 46200
cgcgagggcc ccggcggcaa ggagctggtc tgctacgccg tgccgcgggt ggagagctcg 46260
gagcagggct cgcagcagga gcagcggctg gtacacgagt gggagtccgt gttcgacggg 46320
cacatgtacc gcgaagcccc cgtgggcgga gagccgacct tcaacatcgt cggctggaag 46380
agcagctaca ccggccaacc cgtggccgtg gaggagatga gggactggct gcgccaccgg 46440
gtggagcgcg tgcgcgggct caggccacga cggatcctcg aagtgggctg tggtagggg 46500
ctgatgctct ttgccctgct gccgcactgc gagcgctacg tgggcacgga tttctccccg 46560
gcggcggttg actacgtgcg tcgtacctc cccccggagc acccgggccg cgtggagctg 46620
ctacaccgca cggcggacga gtggagcggc gtggcggccg gtccttcga cgcggtgctg 46680
ctcaattcgg tgggtgcagta cttcccctcg caggagtacc tgcggcaggt gctggccccg 46740

26/38

tgcgtggagg	ccgtggagga	cgggggcttc	gtcttcgtgg	gggatgtgcg	gagcttgccg	46800
ctgctggagt	ccttccacgc	ctcgggtggag	ctggagcggg	ccgccccctc	catgccgctg	46860
gaggcgtggc	gcgagcgggt	gcggcgtgcg	gtgctggagg	acaacgagct	ggtggtggac	46920
ccggcactct	tcgtggcgct	ggcccatcag	catccacggg	tgagccacgt	ggatatcgag	46980
ctgacgcgcg	gcacgcaccc	gaatgagatg	gcacgcttcc	gctacaacgc	cgtgctccac	47040
atcgggcccgc	gcacgccgcc	accggcctcc	gaggtgccgt	gggtggactg	gtccacgcaa	47100
ggactgagcc	tcgatgcact	gcgggctcgg	ctccggcagg	gcccgccagg	gccactgggg	47160
gtggcaggca	ttcccaatgc	gcgagtgtg	cccgcggtcc	gggcgcgcga	ggcgctgggc	47220
agcacaggca	gcgcgcggcg	ggtggaggag	ctgcgccggc	gcctgtccca	acccggcaca	47280
ggggcccagg	atccagaccc	cttctggcaa	ctggccgagt	cgctcggcta	cacggcggca	47340
gtgagctggt	ctcctgggcg	gagggacggg	gccttcgacg	tgctcttctc	cccggcgacg	47400
ccagggatgc	acccgcggtg	gctcggcccc	acgccgctca	accgcacgcc	gccccagcc	47460
tccgcgcgtc	tgtcctcgga	gcctcggcgg	gccagcctct	cactgaggct	gggtccgcg	47520
ctgcgggccc	acctgcagac	ccatctacca	gacttcattg	tgccctcacg	cttcgtggtg	47580
ctgcagtccc	tgccgctcac	tcccaatggc	aaggtggacc	gcgcggcctt	gcccgtcccc	47640
gactcgcgtc	ggctggaatc	tgccccgctg	gtgcccccca	gcaacgagct	ggagcgggtg	47700
ctggcgcagg	tgtggaagga	agtgtctgga	ctggaggaag	tgagccggga	ggacaacttc	47760
ttcgacgtgg	gagggcactc	gctgctgctg	gcgcagggtg	gcagccggct	ggaggccccga	47820
ctgggcccgg	ggctcgaact	ggtgaccctc	ttccgctact	cctccattgc	tgccctggcc	47880
gagcacctgc	aggcccccca	ggagctggca	gccgccgagg	cgcaggtgca	gcgcatggcc	47940
acggagcgcg	ccttgctcca	gcaacaggca	gcccacgcc	gccgggcccg	ttcgaagcgg	48000
ggagctccca	atgacaccta	gtggggatga	agcgtgcag	tcttccatcg	ccctggtagg	48060
catggccggc	cgcttcccgg	gggccccgga	cgtggagtcc	ttctggcgca	atctggtggc	48120
cggagtggag	tccatttctc	tcttctccga	ggaagagctg	agacaggcgg	gcgtgtccga	48180
gcagatacgt	cggcgccccg	agtacgtgcc	cgccaagggt	gtgctggaag	acctggagct	48240
gttcgacgcg	ggcttcttcg	gctactcgcc	gcgcgaggcg	agccacctgg	atccgcagca	48300
acgcctgctg	ctggagtgtc	cgtgggaggc	gctggaggat	gcgggccttc	gcccggatca	48360
gctcccaggc	tgggtgggcg	tctacgtggg	cgccggggac	accagctacc	gcttccagct	48420
gctgcgcgga	cacggggacc	cgctcagcgg	ctccaaggat	gtggcgggct	tctttggcaa	48480
ctaccgggac	tttctcgcca	cgcgcggtgg	ctacaagctc	aatctgcgcg	gccccgcgct	48540
cggcatccac	accgcgtgct	ccacctcgct	cgtctccatc	aacatggcgt	gcagcgcgct	48600

gcgcggcttc gagtgtgaca tggcgctcgc gggcgggctc tcgctgcggc tgcccgcgcg 48660
ctccggctac ctctacgagg aggggggagt cgcctcgaag gacgggcact gccgcccctt 48720
cgacgcgcgt gctaccggca ccgtcacggg ggatggcgtg ggcgtggtgg tgctcaagcg 48780
cctggaggac ggcgtgaagg ccagggatcc catccacgcg gtgatccgcg gctgggcgct 48840
caacaacgac ggcgcctccc gggctggctt caccgcgccc agcgtggagg gccaatccga 48900
ggtgatcgcc ctggcgcacg ccgcggcggg catcagcgca cgcgacatca cctacgtgga 48960
ggctcacggc accggcactc ccctgggaga cccgatcgaa gtggcgggcg tcacccgcgc 49020
cttcggggcg cataccgctg acaccgcctt ctgtaccctc ggtgcggtca aatccaacat 49080
cggacacctg gatgcggcgg caggagtggc cggggtgatc aagacggtgc aggccctgcg 49140
ccaccggctc atccctccca cgctccactt cgagcggccc aatccgcgcg tccacctgga 49200
gcagagcccc ttcttcgtca acaccagcc gctcccgtgg gagagcccgc gcgggcctcg 49260
gctggcgggg gtgagctcct ttggcattgg cggcaccaac gccacacgc tcttcgaaga 49320
ggctccgccc ccgcccgcca gcggccccac ccgtcccaac caggtgctgc tgctctcggc 49380
ccgaagcacc agcgcgctcg agcacatcg aggacggctc gcggcacacc tgcgcgcgca 49440
cccggacctg gaactggcgg acgtggcctt taccctccag gtggggcgag cccgcttccc 49500
ctaccggcgc gccctcacct gtgcacccct ggccgaggcc atggagcgcc tggaggctcc 49560
cgagccgcgg ccacccgagc cgcttgctca cgagggagag cggccccctc tcgtcatgct 49620
ctttccaggg cagggcactc cgctggctcg cacggcccgt gcgctccatg aatccgagcc 49680
caccttccgc caggcgggtg agcagtgtgc ccggtcctg aggcagacgc tggggctgga 49740
tgtcaggag gtgctcttcc cctccgcgga gcaggaggag caggcacgcc ggctggcggc 49800
gcagacgcgg gtggcccagc ccgcgctctt caccctggag tatgccctgg cgcagacctg 49860
gctgggctgg gggctccaac cgcaggcgct ggcaggccat agcctgggcg agctggtggc 49920
ggcgtgcctc gcgggtgtct tctccctgga agatgcactt cagctcgtgg cagcccgggg 49980
acagctcatg cagggatgcc caccgggggc catgctggcc gtgccgctgc cggaggccga 50040
gctggcggcg ctgctgggca gcgagctgtg tatcgccgcg gtgaacgggc cacgggcctg 50100
cgtggcctcg ggtcctctac cggcggtaga ggcgctcacc gccgcgctgg agagccgggg 50160
cgtgtccagc cggcgctgga agacttccca tgccttccac tctgcctcca tggaggcatg 50220
ccaggggcca ctcaccacgc tgctgcggcg catgcgctg caggccccctc gcctcccctg 50280
tgtctccggc ctcaccggcc ggtggctcac cggcgaggag gccacggagc ctacgtactg 50340
ggcccgcag ctccgcgagc cggtagcttt ctccgaggcc ctcgagacgc tctggagcct 50400
caaggagcca gtgctgctgg aggtaggccc cggcaccacg ctcaccgcgc ttgcgcggcg 50460
gcaccccacc cgccccgccc gcacccagga ggtggcgagc ctteccgtgc aaccggacac 50520

28/38

ggcgtgccc	tgattgaag	aggcgtggg	tgagctgtg	caggccggcc	tgagctgga	50580
ctggagcgcc	ctgcacgccc	cgccacgcca	ccgcgcacac	cttcaccct	accccttcga	50640
gcgcagcgc	tactggattg	aaccggaggc	tgccccgcag	ccgcgcgcgc	agcagccgac	50700
gcccgcgtcc	cttgtgcccc	cggagcagcc	ctcgcgcgag	gccctggaag	actggttcta	50760
cgtgcccacc	tgaggagcag	cacctgccac	gagcgggggg	ggacagcccc	tggcggggccc	50820
ggtgctcgcc	ttcatggact	cctcgggcct	ggccgagcag	gtgctggccg	cgctgtggcc	50880
cgcggaactcc	ggcgcgctcc	tcaccgcgt	cgagccggcc	gggcactatg	agcaactcag	50940
tgagcagccc	ttcgcctcc	gccccgagag	cgaggaggac	tgggacgcgc	tcttcaggcc	51000
gctccagtc	cagggccggc	tccccgtcg	catectccat	gcctgggctc	tcaccgcgga	51060
gcccggcccc	tgtacccccg	acggggaagc	ggtgctggaa	cagggcttct	tcagcctgct	51120
gcggctggcc	cgtgcgctcg	gccgccacgc	gcctgagcgc	cctgtacagc	tcgaggtgct	51180
gtccagcttc	gtccacgcgg	tgggcccgcg	ggagccgctc	gagccgctca	aggccacgct	51240
gctggggggcc	tgtgcggtac	ttcccctgga	gtaccctcac	gtgcagtgcc	gcaccattga	51300
cgtgcggccc	gggagcgagc	cacgggaagt	gctggtgagg	agcctggccg	ccgagctggc	51360
cgctcccatg	ggagagagcc	cggtggcctg	gcgcgacggg	cagcggtagc	tgccgcgcgc	51420
cacccggcag	aggctcgagg	cctcgcggcc	cctgcgcagc	cttcgggagc	gaggcgtgta	51480
cctggtagca	ggagggctgg	gtggaatcgg	cctggtgctg	gcgcgagcgc	tggcccagcg	51540
ggcgcgtgct	cggctggcgc	tactcaccca	ctcgccttc	cctccccgag	agcagtggga	51600
gcagtggctg	gaggaagccc	cggcgcaccc	ggagccggcg	tggcggagcg	aggccgaccc	51660
ctcggagcgc	cgcaggaccc	agcaccgcat	ccgtgcctg	ctggagctgg	agcagetcgg	51720
cgccgaggtg	caggtgtaca	cggcggacgt	ggcggaggag	gccgccgtgc	gctcgggtgt	51780
ggagcaggtg	cacgcccgt	gggggaagat	ccacggcgtg	ctgcacgccg	ccgccacgtt	51840
cgacgacggc	gtcatccagc	tcgcacgca	cgagcagtcc	tcgcggggcc	tgcgacacaa	51900
ggtgcggggc	agcatggtgc	tgacagaggt	gctggcgagc	gaggggttg	attggttcgc	51960
cctgtgctcc	tcgctggcgt	cggcgtggg	ctcattcgcc	caggcggact	actgcgcggc	52020
caatgccttc	caggatgcgt	atgcgcacca	cctgcgccgg	cagggcttca	cgggagcgct	52080
ggcgtggac	tggggcacgt	ggagagatac	gggggcggcc	atgcggctgg	tggcacgcac	52140
ccgacggggg	ggccatgaga	agccgcccac	gccgctcacc	caccactct	tcgattgcga	52200
acagcgcgag	ccgggcggga	cgcactggct	cggcctcacg	ctgaggggtg	gagaggactg	52260
ggtggtggac	gagcaccggc	tgcaaggggt	gccaacactg	cccgggggtg	cctacctgga	52320
gctggcgcg	gcggcgtgtg	cccaggcgct	gggggcccag	gcggtggagc	tggcggagct	52380

gttgctgctg gagcctctga cggtaacctg aggcgaatcc aggcagggtcc ggggtggtgct 52440
ccagcccagag gggcagggtc atgccctgcg ggtggagagc cggtcggagg aagcgcgggg 52500
atggaatgag catgcgcggg gccgggtgcg cggcgtgcct cggctggccg agcgcatcca 52560
gcccagagctg ctgcgcgccc cctgcgaaca cgagcagccc gtgcccggag agccccagga 52620
gcaaggccct gtccacgccc gagcacgctg gcatggcctc ttccagtggg ttccgcccgg 52680
ccctcgccag gccctcgccc agctcgcgct accggagccc ttccacggtg acctcgagcg 52740
cttcgagctc caccggcgcg tcatggacat ggccaccagc ttccgccattc ccggcggagt 52800
gccctggctc gcctttggct atgagcgcggt gctcattcac ggtccgctgc caccgcagggt 52860
gctcagccac gtgagcctgc ccgaggagtc gcaggctggc gcgcaacaac tccggctaca 52920
gggtcgggctg ctggatctgg agggctggga gcgggtgchg atcgatgggt acctgctgcg 52980
gccgctgaag cccagcgacg ccagtgtcga gccggcgggc cccaacgtgg aagtcgcccgt 53040
ggggacgccc gggctgctgg agagccttgg cctgcgcgcg tgcaaccgcc ccgcccccg 53100
gccgcgccag gtggagatcg aggtggaggc cgctgggctc aacttcctgg acgtgctggg 53160
cgcgctgggg atgatgccgg cattggaggc ggaggagagc gtactggggc gcgagtgtc 53220
cggacgcatt gccgccgtgg gcgagggcgt cagcgggctg cgcggtgggg acgaggtgct 53280
ggcggtggcc ccaggctgct tccgctccta cgtgctgggt gatgagagcc aggtggtgch 53340
caggccccgc tcgctggggc tcgcccaggg ggcgggccag atggtgccgt tcgccacggc 53400
gtacttcgcc ctgcacaccg tgggcccgtt gcggcgcggc gagcgcatcc tcatccacgc 53460
cgcgggccga gggctgggtc tggccgccgt tcagctcgca tcccggaccg gggcggagat 53520
attggccacc gcgggcagtg agcagaagcg cgagtacctg cgctcgctcg gcattgctca 53580
cgtgctggac tcgcgcagca cctccttcgt cagcgaagtg cgcgagcgca ccggcggggc 53640
tggggtggat gtggtgctca actcgctggc gggagagctg ctccctggcg gcctgtccgt 53700
gctggccccg cacggccgct tcctggagct ggggaagagg gacctgtatg cggaccagca 53760
gggtgggcctc cgaaccctgg cccgagggca gactttcgcc gccattgact tcggtcccca 53820
ccaccggac ttccgagcgg tgctcgagga ggtggccacg caactcacc agggccagct 53880
cgagccattg cccaccgcc tcttccccgc gcggcagggt gccgaagcct tcagcttcat 53940
ggccccgcgc ctgcacatcg gtgcgctgc cgtctccatg cagggggcga cggcattgcc 54000
cgcgctccatg actcggggct ccaggcccg accggtggcc gtacctcctt gggaggaccc 54060
gcggctggcg ggcggcatct cctctgaaga gggcgcgag gccttcctgc gcgcgttgga 54120
gcaaggggca ccgcagctca tcatctcccc ccaggacttc agctcgctgc tgcgcggcct 54180
gggcggcagc cagggcgtgc gcgaaaagga gcgcctcgtc accgggcgcg ccgcccgcgc 54240
cgagccgcag gccttgccac cctcctcgct ggagcagctc atcgagcagg tgtggcgcaa 54300

30/38

gcacctgggt gtggagcgcg tgcagccac ggacagcttc ttccagcttg gaggagactc 54360
gctgctgggc atccaagtgg cggcggatct ccgcaggcac ctgggtgtgg agctgcccac 54420
ggccaccctc ttcagccacc ccaccctcgc cgcgctggcc gcggctctgc gggctcgaca 54480
gggtgaggcc gcggctccca ctgccccgc gccagcgctc gtgccggatc cagccgcgcg 54540
cttcgagcct tccccctca cggatgtgca ggaagcctac tgggtgggtc gccgctcggc 54600
cttcgagctc ggcggcgtcg ccgcccattg ctacttcgaa atcgaaagcc cggggctgga 54660
ggtggagcgc ttcattcaat gctggcgcca gtcctgcag cgccacgaca tgctgcgcatt 54720
ggtggtgctc ccggatgggc ggcagcaggc gctcgagcag gtgccggagt acacgccgga 54780
ggtggtggag ctgcgggggc tctccccca ggaggccgag tcccgccggc tccagctgcg 54840
tgagcgcatg gccaccagg tctgcgcgag cgaccgctgg ccgctcttcg agctggtgct 54900
ctgccggtac gagggaggcg tccgcatcca catgagcatg gatgccctga tgctggatgc 54960
gtggagttca gcggtgcttc ggcaggactt cgcacagctg taccacgagc cgggccggcc 55020
gctggagccg ctggccatca ccttcgcgca ctacgtgctg gcggagcgcc ggctgcgcga 55080
gggcgaggcc catgagcgcg ccgcgcata ctggtgggct cggctggaca cgctgccgcc 55140
accgcccag agctcccctg tgaaggaacc ctgcagctg gagcacgcgc ggttcaccca 55200
ccgcgaggct cggctcgagc cacaccgctg ggcccggctc caggagcggg cgcgcgccca 55260
cggcctcacc cctcgggccg cctgcatggc cgccttcgcc gaggtgcttg cccgctggag 55320
ccgtcaccgc cgcttcaccc tcaacctcac cctcttcag cgcttgcccc tgcaccgcga 55380
ggtggacgag ctggtgggcg acttcacctc cctggtcctg ctggaggtag aggcacacgc 55440
ggcgagcacc ttcgcccagc gtgcctcccg gctccaggca cagctatggc gggacctgga 55500
gcacggcagc gtgagcgccg tgcagctcat ccgcgagctc gtccgcaccg gccgcgcctc 55560
cccgggcgcc atcatgcccg tcgtcttcac cagcatcctc agcctggatg cgcggcgcg 55620
ccccagggc agcctctcct tcttcgaggg agaactggtg tacagcatca gccagactcc 55680
ccaggtgtgg ctggaccacg gagtccacga ggaggagggg gcgctcgtcc tggcgtggga 55740
ctcgggtggag gcgctcttc ctcggggcat ggtggacgac atgttcacg cctaccagcg 55800
gctgctgggg gcaactcgcc aggaggagca agcgtgggag ggcgagctgc cggagctgct 55860
gcctcctgcc cagcgtgagt tgctcgcgcg ctacaacgcc accaggcgc cgcggcccag 55920
cggacggctg gaggagggt tcttcaccca ggcgcggctc caccgcgagc tccccgcgt 55980
gctcgcaccg gagcgaccc tgagctacgg cgagctggca aggcgagccc aggcgctggc 56040
cgcgcgccca cgcgagctgg aggtgcagcc tcaggagttg gtggccattg ccatgcacaa 56100
gggctgggag caggccaccg ccgtgctcgg cgtgctgcag gcggccgcgg cctacctgcc 56160

gttgatcca gagcagccgc cgctccggct ccaccagctc ctggaggagg ggcccggccg- 56220
cgtggtgctc acccagtcct ccttgctgca caccgtgccc tggccggccg gtgtgcaggt 56280
gatcgccgtg gacgaactcg agcctgcaac cgaggccccc ccgctaccgc ctgcggtac 56340
accggagcac ctgcctacg tcatctatac ctccggctcc accggcaaac ccaagggtgt 56400
ggcattgaa caccgggccc cgctcaacac cgtggtggac ctcaacaccc gctttggcgt 56460
cgcccggag gatcggtgc tgggcctgtc cgcgtcacc ttcgacctgt cgggtgtacga 56520
cgtgctgggg ctgctcggcg cgggcggagc gctggtgctg cccgcggcgg aggcggagaa 56580
ggatcccgcc cactggtggg agcggctggt ggctgggcgg gtcacggtgt ggaactccac 56640
cccggcgctg atgctgctgc tgggtggagta cgccgagcag cgccggctga agctgcccgc 56700
ggcgtgctgc ctggtcatgc tcagcggcga ctggatcccc gtggcgctac cggatcgcat 56760
ccgcgcactc ggcagggacg tgcagggtgt gagcctggga ggcgccaccg aggttccat 56820
ctggtccatt gctacccca ttggccaggt ggcaccgcag tggagagca ttcctatgg 56880
catgccgctg gccaatcagc gattccacgt gctggacggg cggctggagg cacggccctg 56940
gtgggtgccg ggcgagctgt acattggcgg cgaggggctg gcacgcgagt actggaggga 57000
tgagccgctc accgcaacgc gcttcatccg tcaccgcgc acaggcgagc ggctgtaccg 57060
caccgggga ccaagggcga tgctgcccga agggagcatc gattcctcg gccgggagga 57120
tctgcaggtc aagggtgcagg gcttccgcgt ggagctgggg gagattgaag cggcgctggc 57180
ccagcaccgc gcgctatcg cgagcgtggt ggtggcacga ggagagcctc ggggagtgcg 57240
gcggctggtg gctacgcgg ttcctcgctc ggtcagacg cccgcggcag gcgagctgcg 57300
ccgctacctg gcggagcggc taccgcgta catggtgccc tccgccttcg tctcctgga 57360
gtccctgccc cgcagccga acgggaagat cgcccgagat cagctccccg agccgcagca 57420
gacacaggga ctggtgcac aggcgcggc ggccgacccg ctgctggagc ggctggcagc 57480
cctggtgaaa gaggcgtga ggctcgagcg cgtggagccc caggacagcc tgctggacct 57540
gggggcggac tcggtggcg tcatccgcct catcaaccgc ctggaggcag agctgcagtt 57600
ccgcccgcga ctggccgaca tctacgagaa ccccaccgtg caggggctcg ccacgtcca 57660
ccaagagaag acaaagagcc agggagagg aggcgctccg cggctcacgg cgcgccgctc 57720
cacgcttctg cccgccgagg agtgggggcg cttcaaggcc aaccgcccgg gcctgcgcg 57780
cttccccgat ggaacgctg aggtggcact gccaggcagc ggcttggcg cggccccga 57840
ggagctgaca gctctgagc gccgcgcag tgtgcgcacc tactccctcg agcccgtag 57900
ccatgagcag ttgggacggc tgctggcccc cttgcgcgag tgggaggtgc agggctcgcg 57960
gcgttacctc tacgctccg cgggcgggct ctaccgggtg cagctctacc tccacctcaa 58020
gcccggccgg gcgcgagggc tcgagcccgg cacctggtat tacgatccca gcacccatcg 58080

32/38

gctggtgctc ctgtccgccc gagccgggct ggatcgccgc atccatgatc cgcaccagaa 58140
ccaggccatc ttcgactcgg ccgcgttctc cctcttctc atcgcccga tgggggcccgt 58200
ggagccggta tacgccgaac atgcctgca cttegccacg ctggaggcag ggctgatgac 58260
ccagcttctc gatctggggc cggctcccag tgggctgggg ctgtgccata tcggcgacct 58320
ggacttcgcc caggcgccgg ggctcttcca ctggaggag gagcacgtgc tgctgcacag 58380
cctggtgggc ggagtactcc cactcgggg gcaggaggcg gcctcgggtc ccgctgaggg 58440
aggaacggag gcgcggcaac tggcgagct cctgcaacag gtgaagacgc tcacgcccga 58500
ggccgcgcga gcgctgctcg aggccgccg cggcagcaag gggagaccac atgagtaagc 58560
tgcacgagga gctggagagc ctggcgccgg agcagcgga gctactcgc gccctgatga 58620
aagagcaggg cctggacgag ggcgcattgc tgatgcccg ggagcgcaag cccgaggggc 58680
tgccgctctc gtccgcccag cagcgcatgt ggttctgca acagctcagg cctacgagcc 58740
ccttctacaa cgtgcacgcg gcgctccggc tcacgggaaa gctgaagggtg gagtgcctgg 58800
tgcacagcct gaatgagttc gtccggcgtc acgagccgct gcgcacgggtg tccccctccg 58860
ccggaggtca acccctgcag cgcctcctcg cgcgcgacc ggctgcgctc gagcagcgcg 58920
acctgagcgg cgtgcccgcc caggagcgag aggcgaggt gtaccgcgcg gtggagcacg 58980
cagcgctcgc ctctttgat ctggagcgag agccgccgtg ccgcttctc ctggtgagcg 59040
tgagccgga ggaacacgtg ctggtgttcg ccacgcacca catcgagcg gatggctggt 59100
cgctcggcgt ctctgtgag gagctgtgcg cgctgtacac ggctgctgta cacgcggagc 59160
caccggcgct gccaccgctg cggctgcagt acgcggactt cgcgcggtg gagcggagcc 59220
ggctcaaggg tgggcgcgag cgcgagctgc tcgagtactg gcaggagcag ctggcagggc 59280
tgccggacct gagcaccctg ccgccagagc ggcccaggcc cccgctgtcc aagcaggagg 59340
gcgccacgtt cgaattcgcc ctgccaccac accaggtgca ggccctgcgc tcacttgctc 59400
aagcgcgcg cactcgcctc ttctccgtgc tgctcgccgc cttccaatgg ctgctggccc 59460
ggtgcgcccg ccaggacgac gtggccctgg gcatgcccac cgccaaccgc aaccgcaagg 59520
agctggaagg cctcatcgcc tgcttcgcca gtaccctggt gctgcgcgcg aagccctccg 59580
cctcgaccgc cttcacctca tggctcgccc aggtgtccga gcagctccac ggcgcctcg 59640
agcaccagga ggtgcctttc gaacgattgg tggaggtgct ccagccccgg cgcgcgatgg 59700
atcggcaccc gctcttcag atcttctcgc ccatgcagca gcaccgctg cggcgcgccg 59760
agctgccagg cctgctgctc tcggagttcc cccttcgcag ccgctgggcc cgcttcgac 59820
tggagttcca cctctgggag tctgccagg gcgtcgaggg caccctcatc tacgacgtgg 59880
acctgtatgg ccaggagggc gtagcgcaac tgggtcgcgcg ctacgtgagc ctgctcgaag 59940

ccgtggcggc ggacccgacg cggacactgg gggagctggc gggtagagacg ctggccccgc 60000
ccgtacggaa gcaggttctc gccctgtcgg agtgtccgcc cctcccacgc ccgagcacag 60060
cccctcggac actggccgag gcacttctac agaccgccga gcgcttcccc acggccacgg 60120
tcagcttcgt ccaggcacag ggctcttgca cggcctggac tctgccagag ctggtggagc 60180
gggcccggcg cctgcaggcc ggcttgccgc agtggggatt gcggccgggc gacagtctgg 60240
tgctgggtgct ggggagggag gaagagacgg tagaggcact ctgggcctgc gtgctggcgg 60300
gagtggcgcc cctgggtgcta ccggcgccgc ccgcgagagc ggaggcgagt cccgcgctct 60360
cacgtctgcg gcatgcccgg cagctgctgg gtgggccccg ggtcctcacc cggcaggaga 60420
tgctgcccga tcttgccagg cagcttcagg tctcgcccac cgccgacatc ctggggggccg 60480
tgaggagagct gagagccacg ggtggcgagg cgccctgcc ccaggggcg atggatgacg 60540
tgccctgct caacctcacc tcgggcacca ccggcaaggc caagtgcgc atgtcacgc 60600
accgcaacct gctgggtgcg ctggaggcca ccaacgtcgt ctacgagtcc cagccgctcg 60660
agcgcgccct cgtctggttg cagctgcaca acattggcgc cctgtccgag taccacctcc 60720
gcccgtctcg cgccggcatg cacaccttc acgtccac cgaggagggtg ctgcccagac 60780
ccctgcgatg gctggagtgg ctcgagcgg atggcattgc acagacgtgg gcgccagct 60840
tcgcctacag ccacctgctg gacggctgc gcaagggtga ggaccggcg tggagcctgg 60900
gaggcgtgcg cgtgctgctc agcgccggag agcagatctc cgcgcccatg gtggaggagc 60960
tgatgcggag gttggctccc tcgggagtgc gggaggatgc ctctgctgct gcctggggaa 61020
tgaccgagac ggcttcgggg gtgacctatg cgcggcggcc cggaacgccg ccccgcatgc 61080
acacctgga gcgcgcgagc ttgagcggtc cgcttcggca cgcggcacc gcttcgcccc 61140
ccgcgtgcg gctgatggat gtaggagccc ccattgccgg caccgccctg cgggtgggtg 61200
acgcgagcgg cgagctgctc tccaggagt gcgtggggcg cattcaagtg cgcgccgaga 61260
tgatctctcc cggtactac ggggacccga aggcctcggc ggcgctgctc accgccgacg 61320
gctggctgga gacaggcgat ctgggcttcc tctccgagg agcgctgacc attaccggcc 61380
gcgccaagga tctggtcatc atccacggca ccaacttctc ctgctacgag atcgagtccg 61440
cggtaggca ggtggaggga gtggcgccct cctcggcggc ggcgggcggtg gtgcggatgc 61500
tcgaggggag caggaggag ctggcggtct tcttcgtgcc caccggaggga ctgctccgc 61560
aaccctcgc ctactcctg tcgcgcatcc gccagcaggt gctcgagcag gtgggggtgc 61620
ggatcgatca tctcatccc ctcgagctc accagctccc acgcacggag ggaggcaagc 61680
tgcgccgctc ggagctgagg gcgcgcttcg aagcaggtga gctgaggcg ccgcaaccg 61740
ctccggctcc gagccctcg cggcgctgg agcagctcat cgctccgctc tggcgggagg 61800
tgctcgagca ccaggacatc gcgcccagg ccagcttctt cgacctgggc ggcaactcca 61860

34/38

tcctcctggt gcgcgtggag cgggcactgc gcgcgcggct cggcctggag ctgacgtca 61920
tggacctgtt cgcctacccc acggtgcact cgctggcgga ctacctggag cctcgagccg 61980
cccagctgcc agcacaggcc tccccctcca cgcaggccga gcgccggcgg gggatgcgcg 62040
gcgcggcgct ggagcagaga cgcactcgcc gccaggcaca acgggacgag gactagacac 62100
gcatgagcca cgacacgcag ctcgagggag ccgtagcggt agtggccatg gcggggcgct 62160
tccccaggc cccaccctg gaggactact ggcgcatcat ccacgagggc gtggacgcct 62220
ttaccagct cacggacgaa cagctcctgg cctccggagt ggggccctcg ctctgcgcc 62280
aaccgcgtta cgtgcggcgc gccgcctgc tggacaggat ggaccagttc gacgccgcgt 62340
tctttggctt ctccccgca gaggcggagg tgctcgaccc acagcacggc ctcttcctgg 62400
agtgcgcgca cgaggcgctg gagcgggccc gccatggctc cgagcgcgca agggggcgcg 62460
tgggggtctt cgccagcgcc agcctcaaca gctactacct gcattccctg cacgggaacg 62520
cccggctgag ggaggtgctg ggagacttcc agctcgccat cgccaacgac aaggacttcc 62580
tgccacccc cgtctcgtac aagctgggcc tgcgcgggcc cagcgtggcg gtgcagacgg 62640
catgctccag ctcgctggtg gccgtccacc tcgcgtgcca gagcctgctc aatggtgagt 62700
gcgccctggc gctcgcgggc ggctcctctc tctccgtgcc ccaggcccag ggctacctct 62760
accaggaggg aggcattgcc tcaccggacg ggtactgccg ccccttcgac gcggcggcgg 62820
cgggcaccaa ccgaggcaac ggcgtggggg tgggtgctgct caagctgctc gaggaggcgc 62880
tggcggacgg agacaccatc cacgccgtca tccgcggctc ggcggtgaac aacgacgggg 62940
cacacaagat cggctacacc gcgccaagcg tggaggggca ggcctcggtc atctccgagg 63000
cgctggaggt ggcggccgtc tccgcggaca gcatcggtc cgtggaggca cacggcaccg 63060
ccaggccct gggagatccg atcgaggtcc aggcctcac gcgggccttc cgcaagcaca 63120
cggagcgccg cggctactgc gctctcggt cggtaaggc caacatcgga cacctggatg 63180
cggcagcggg catcgccagc ctgatcaagg ccgtgctggt gctggagcgc cgccagctcc 63240
cgccatgcc gcacttcacc tcccccaacc cacgcacga cttcgagcgc agccccttct 63300
acgtgagcgg gcgggggag cegtgggagc cggtggaagg cccgcggcgc gcgggcgtca 63360
gtccttcgg tattggtggc accaactgc atgtggtgct ggaggaggcg ccgccccggc 63420
cggcgcggcg gcgggagcg cgccctacc agctcctgcc gctgtcagca cgacaggagc 63480
gcgcccggga ggaggtcgag gcccggtgc gcgagcacct gcgccagcag cccaggagc 63540
cgctggagga ggtggccac acgctccagg taggcgcgg ccacatggcc tggcgcacag 63600
cgctggtgag cagcagccct gccaggcgg tagagctgct ggagaggcgc cgaccgagg 63660
cgctgctgcg tggccagagt gcggcccagg cccgctccgt ggccttcctt ctaccgggcc 63720

agggctcaca gtacgtgggc atgggagcgg cgctccatga gtccgaggcg cccttccgtg 63780
 aacaggtgga tctctgtgcc gggaaactcc agggcgtgct cgggttggac gtgcgccgac 63840
 tgctgtaccc tggcccgcgg gagcgggaat gggcccagga gcgcctgcgg gagacgcggg 63900
 tatgccagcc ggtgctcttc accgtggagt acgcgctggc ccgactgctg gaggcctggg 63960
 ggatacgcgc caccgcgctg ttggggccaca gcctgggcga atatgtggcc gcgtgcctcg 64020
 cgggcgtggt ctccttgaa gaggcgctgg aggtggtggc cgcacgtgga cagctcatgg 64080
 gcagcatgcc cccggggggc atgctcgccg tggggctccc agcccaggaa gtggaaccgc 64140
 tectaccgc ggcactggcc ctggccgcgc acaacagccc ccaggcctgt gtggtggcgg 64200
 gcccaccga gccctggcc gagctgcgcg cagcactga gcagcggggc atcgcgtgca 64260
 cgcctctggc tgtctctcac gccttccact cgccgatgat ggagccggcg gtggcccat 64320
 tcgtcggcg gctgcgccc atgcatctgc ggcctccgag cctgcccttc atctccaatg 64380
 tgactggcac ctggatagag gccgaggagg cgacctacc ggagtactgg ggccggcacc 64440
 tgctgcagcc ggtgcgcttt gccaggggc tggagcgggt gtgcgagggg gtgcagccgc 64500
 ggctgctggt ggaagtgggg ccggggccaca cactggggcg gctggcgggc cgccaaagct 64560
 ccggccccgt gcaggtggtg tccacgctgg gctccagccg ggaggaaagc tcggaggtgg 64620
 agcggctgct gacggcaatc gggcgactgt ggggtggagg agccgaggtg gactgggctg 64680
 ggctccaccg cggcgaacgc cggaggcgcg cggtgctacc cacctaccct tttgagcacc 64740
 agcgctactg ggtggagaca gccccccagc cttccccctc cgagcggcct gggacacctg 64800
 ccgagagccc tgtcaccagc agcttctacg tgccaggttg gagccgcgcg gctctgccct 64860
 cggcgcccc ctcgagacgg gtggggccggc tgctgttgtt ggcgagggcc catgggtgga 64920
 gccaggggct cgcggagcgg ctccggggcc ggggacactc cgtcacctc gtcgagccag 64980
 gtgagcggct cgagcgctg acccgggagc actggcgctt gccccgggc cgcgcgagg 65040
 acttccaacg cctgctggag gactccggag agctgccctc gaggtgctc cacctctggc 65100
 tgctggggag cagccagagc ccccaggagc gcggcttcta caccctgctg gcgctggccc 65160
 aggcactggg agcccacggc acccggcccg cgggtggagct cacggtggtg accgatcagc 65220
 tctacgcggt ggcagagggc gagccgatac agccgctcaa ggcgctcctg cagggccctg 65280
 ctccgtgct tccccaggaa ctccccgggt gcacgtgccg cctcgtggac atggccctgc 65340
 cgccaggcgg cgtggcggag gagctgctgg agcggctggt ggcggaggtg gagtccaccg 65400
 cgtccgaacg ctcggtggcg taccgcggcg cggctcgtg ggtgcgcgag ttcgtccccg 65460
 tgccgtgcc cccggccgag cttcgcagc tcccgttcg gcagagaggg gtgtatttga 65520
 ttgtcggcgg cctggggggc gtagggctcg cgctggcaga gcacctggcg cggcgcgtgt 65580
 cggcgcggt ggtgctcacg ggccgctcgc ccacgccacc gcgcgagagc tggagcgcac 65640

ggctgggcac	tcccacgcgg	ctgcggctct	cacaggagct	cgagtggctg	cgcggtgccg	65700
ccgagcagat	cgagcaccag	cggccccctgc	actccctggc	cgacagtccc	ggcctggaag	65760
agtcgctgcg	cctgctgtgc	gccagctacc	tctaccactt	tctcttccct	cttcaggcgc	65820
ctctgagact	tggagagccg	agggcgatgc	aggcattgcg	tgagcgcctg	gggctgcagc	65880
ccggcttcga	ggcctcttc	tccttcatga	tcggcaccct	ggagcaggcg	aagctgattc	65940
gagtgggaga	gggcatgctc	gagggtcagg	tggagccagc	ccacgtgccc	acacctcgag	66000
cactgcacga	gcgactgctc	gaaggctacc	cggaggtgtc	aggcctgctg	gagctgcttg	66060
agcatttgtt	gcgccactac	cctgaggtgc	tccgagggag	cctgccggcc	ctcagtgtgc	66120
tctacccttc	gggccgctcc	gaggaggact	ccgcacgaag	cggggtcgag	tggagcagca	66180
tggggcagtg	cgtggcgctg	ctgcggcatt	tcctcgcccc	ccatgccgag	cgcacgcagg	66240
ggcgtaacct	gcgcatactc	gaggtgggcg	gcggcagtg	cgtgggtgctc	caggccctgc	66300
tgcgctgct	gcggcaccac	ccggtggagt	accacttcac	ggatattggc	ccctccttcg	66360
tccgcgccat	ggagggagtc	ggccgcgcgc	agggcctcac	cttcctccac	acctcggtgc	66420
tggacatctc	ccgccctccc	ccagagcagg	gccatccgcc	tggctcgat	gatctggtga	66480
ttgccctcaa	cgtggtgcac	gccacccgc	gcgttctca	gtccctggcc	cacctcgaga	66540
gcctgctcgt	cccaggcggc	cacctgtgcc	tggtggagac	ggtgaagcag	cagccctggg	66600
tggacatgat	ctggggtttg	gccgaggggt	ggtggagcta	cgaggacgag	ctgcgcaccc	66660
gctctccctt	gctggaggta	ggcgactggg	agcacgccct	gcgcgacgtg	ggctttgccc	66720
aggtggagggt	gctgccagcg	gccgtggagc	agcgctcacg	ctgggacaac	gtgctcctca	66780
ttgcccagcg	gcccggcgag	agcgggcttc	agccagtggg	cgggcgcgcg	gccatgcagg	66840
agcggatccg	ccggctgcgg	gccattgaag	aagcgggagg	cgaggtgctg	cccctggtgg	66900
cggacgtcac	cgatcgggag	cgcattggcg	aggtgctcgc	cgaggtgaag	cgcgcgccatg	66960
gggcgctcca	tggggtgatc	cacgcgcgc	tggtgctgga	ggatgggctg	atgcagctca	67020
agacgcgcga	gtcgccggg	cgcgtgctgg	cctccaaggt	cgagggcacc	ctggtgctgg	67080
acgaactgct	gcgcgacgag	ccactcgact	tcttcgtcct	gtgctcctcg	ctgagcgcac	67140
tgctcggcgc	gctggggcag	gcggactatg	ccgcgcgcag	tgccttcttc	gacgcgtatg	67200
cgcactcaca	gcgagggcgc	acggaccgcc	gcaccatctc	catggactgg	gatcgctggc	67260
tggaggtggg	cgcgcgccatg	cgcctggggc	tggggtggc	ggcgggagca	ctggggctgc	67320
agcgcacggc	acccggagag	tacacggtgc	gatggcaggc	cgagcgcctg	tggtggctgg	67380
acgagcaccg	cctggaagga	cgggccacgc	tgccgggggt	ggcgtaacctg	gagctggtgc	67440
gagccgccct	ggtccaggaa	ctgggcgagg	ccccgtgga	gctggagcaa	ctggtgctgc	67500

tctctctgct	ggaagcgccc	gcgggggagg	aggtggaggt	gcgattccac	ctgcgaccgg	67560
aggatgaggg	ctatgcgctg	gagatccgca	gccgggctgg	cggcctcgcg	aatgggggct	67620
ggcgcagcca	tgcgatgggc	cggttgagga	tgctgcctcg	cgggagcgct	cgacccaccc	67680
acccgctacg	ggagctggag	gaacggctcg	gcttgacca	ggcgccgagg	gagcatgagc	67740
cagctcaggg	cccccgcgag	ctggcacctg	ccctgggccg	gcgctggtcc	tcgctgagct	67800
ggcaccgcgc	gtggaaggga	gaagagggcc	tggccttgat	cgagctcccc	gaggagctgg	67860
ccgaggatct	gcggcaatgg	cccctgcacc	ctgccctgct	ggacgcggcg	accgggtttg	67920
caccgctgcc	ccccggtgcc	tggtcgccgc	tgtcctatgg	caacacccgc	atccacggcc	67980
cgctgccgcg	gcagctgtac	agccacatcc	gccggctgga	gcccgcaggt	gcccaggccg	68040
gagtgggtgcg	gctggaggtg	cgctcatgg	acggcgaagg	gcgggagctg	ctgtgcgtgg	68100
aggagtctgt	gctccgcccg	gtggaggtgg	actcgttggc	tcggccacag	cccgtggca	68160
aggctccgcg	ccaggcagtg	cttccacgac	caggcgctct	cgactcgctg	cgactgcaac	68220
ccctggagcg	ccttcgcccg	caggagggag	aggtggaggt	gcaggtgctc	gccgcgggcc	68280
tcaatttcaa	ggacgcgctg	ctggcgctgg	gggcgttgcc	agtggagctg	gccaacggcg	68340
ccccactggc	gctgggagtc	gagtgcgcgg	gcacatcttc	cgcggtaggt	ccaggagtga	68400
gaggcctgcg	agtgggcgag	gcggtggtgg	ccgcggcagc	aggagccttc	gcctcgcatg	68460
tgcgagtccc	ccaggagcag	gtcttcccca	agccagcggg	cctgagcttc	gagcaggcgg	68520
cgatggtgcc	cgtcactctg	ttcacgcctt	ggtatgccct	ggaggagctg	gcgcggctgc	68580
gcgcgggaga	gcacgtcctc	atccacgccc	cggccacggg	ggtgggactg	gcgggggtga	68640
agttggcgct	gcggcggggc	gccacggtgt	acgccaccgc	gggcagcgag	ccgaagcgcg	68700
agctgctgcg	ctccctgggg	gtaacgctgg	ccatggactc	gcgcgcgccc	gggtttgacg	68760
accagcttct	gcaacacacc	caggggcgag	gcgtggacgt	ggtgctcaac	tcgctgagcg	68820
gcgagttcct	ctcccgagc	ctgggcgtgc	tcgctcgcca	cgggcgcttc	gtcgagctgg	68880
gtgtaaggga	catgctctct	ggcggcacgc	tgccgctggc	tcccttcgag	cgcgggctca	68940
ccttcctcgc	ggcgagata	gacccgggga	tgaagggcta	ccgggagctg	atgggcgagg	69000
ccctgcggca	gatcgagcgc	ggagagctgg	agccgctgcc	ctacacggca	tggccactgg	69060
agcgcgtggc	cgaggcgctc	cagctcgtct	ccaagggcag	gcacgtaggc	aaggtggtgc	69120
tcaccccgga	agagcccttg	tcacggccgc	cagcgtccc	gagctcctca	aggccagtgc	69180
aaccccgggc	aggcgtggcg	attgtcgggc	tgcgctccgc	cgagggtgc	gaggccttcg	69240
agcgcctcct	cgcgcccggc	ctgccgcagg	tggccgtctc	cacccgtgag	ctgcgcgcgc	69300
gcatggcgga	gatcgagcgc	ctgcgcgtct	cctcctggga	gccggcgctt	ccctcggtgc	69360
cccgcggctc	ccccgccagg	accgaacgct	ccaagcccta	tgttgctcct	cgacccgagc	69420

gcgagcgcg cctggccgcg ctctggcagg acctgctcgg cgtctccgag gtgggggttg 69480
 atgatgactt cttcgagctg aaaggggact cgctgctcgc catccagctc atggggcggt 69540
 tgcgcaagga gctggagggtg gaattgccac tctcctcctt cctcgcccgc cctacgtgc 69600
 gcacactgct ctccgccatg ccagccccgg caacggacga gcctgcctcg gccaccgggg 69660
 caatcccagc cgccgctccc cagccgtcac cggctgtctc cagcgaaccc gctcggaagt 69720
 ggcgccacct ggtgcccata caacctcagg gcgagagccc tcccttcttc tgggctgcgc 69780
 cgctgatggg caccgtgttt cctacttca cgctggcccg ctgcctggcg cccaccacc 69840
 ctttctatgc tctccagccg ccagggttgc aggaggggca ggagcccatg gacaagggtg 69900
 aggaactggc ggccctgtac gtgcgcgagc tgagagagct gcagccgcga gggccctacc 69960
 gcctgggagg ctggtccttc ggctgcgtgg tggcttatga ggttgcccta cagctcgagg 70020
 cggctggaga gcagggtggc ctgctgagcc tgctggactt cctgcccct tcgggccagc 70080
 gagccagtct ctggggagcg gcccgcttct tcagcggctc ggtgctgcgc ggcctggcgc 70140
 cctatacgct cgactatctc tacctggcgg ctgcgcctcc tctgaatcca gacgtgcct 70200
 cgcgcctgag cagctccctg acccgaggat gggagcgcct gcggaaagga gggctggtgc 70260
 aggaactgct ggaccatgcc gccatggccc ggggtggtgcc tgaggactca cggctgctgc 70320
 tgctgcgaga gcctggcatt gcccccatgt tgaggctgac cggggtcac cagcgagcca 70380
 tgctctccta tcggccccga ggtgcctcc gccagcgcct cgtcctcttc cgcacctccg 70440
 agcaggcctg gagcttcgcg cgagacctgg gctggggagc cctcagcgc tcgggcgtgg 70500
 atgtgcgca ggcgccggga gaccatatga cgctgctgcg gccacctcac gtggagcggg 70560
 tggcagagca gctccgggcc ctgctcagct ccgcgcctct cccgactcgc cgctgagttc 70620
 gggctgcgcg ccagaggctg cctcatgcat gccctttggc agccagcccg agctgtccac 70680
 ccactccagc gcctcgctca gatcggtgctt ttcgacgata tgatccaagg tgaccacctg 70740
 cgcttccacg ctgttgccca gatgagaggg caccgcgga tc 70782